From the Department of Medicine Solna
Karolinska Institutet, Stockholm, Sweden

FAM167A-BLK IS A SUSCEPTIBILITY LOCUS IN AUTOIMMUNE DISEASES:

CHARACTERIZATION OF THE FAM167 GENE FAMILY

Lara Mentlein

Stockholm 2018

All previously published papers were reproduced with permission from the publisher.
Graphical abstract was generated at www.wordclouds.com by Lara Mentlein.
Published by Karolinska Institutet.
Printed by US-AB.

ISBN 978-91-7831-224-5

FAM167A-BLK is a Susceptibility Locus in Autoimmune Diseases: Characterization of the FAM167 Gene Family

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Lara Mentlein

Principal Supervisor:
Professor Marie Wahren-Herlenius
MD, PhD
Karolinska Institutet
Department of Medicine Solna
Division of Rheumatology

Co-supervisor:
Senior Researcher Alexander Espinosa
PhD
Karolinska Institutet
Department of Medicine Solna
Division of Rheumatology

Opponent:
Associate Professor Kristina Lejon
PhD
Umeå University
Department of Clinical Microbiology
Division of Immunology

Examination Board:
Docent Saedis Saevarsdottir
MD, PhD
Karolinska Institutet
Institute of Environmental Medicine
Division of Cardiovascular Epidemiology

Professor emeritus Thomas Skogh
MD, PhD
Linköpings Universitet
Department of Clinical and Experimental Medicine
Division of Neuro and Inflammation Sciences

Professor Ewa Ehrenborg
MD, PhD
Karolinska Institutet
Department of Medicine Solna
Division of Cardiovascular Medicine

The public defense will take place on Friday 16th of November 2018 at 9:00 am in the CMM lecture hall (L8:00, Karolinska University Hospital Solna).

„Alles Wissen und alle Vermehrung unseres Wissens endet nicht mit einem Schlusspunkt, sondern mit Fragezeichen.“

“All knowledge and all increase of our knowledge do not culminate in an end; instead there will be more questions.”

Hermann Hesse

ABSTRACT

Autoimmune diseases are complex multifactorial diseases, and their pathogenesis is only partially understood, but both genetic, as well as environmental factors, have been correlated to an increased risk for the diseases. Genome-wide association studies of Sjögren's syndrome, systemic lupus erythematosus, and rheumatoid arthritis have identified many risk loci; most studies include associations to the *FAM167A-BLK* locus. A significantly increased expression of *FAM167A* in cells carrying disease-associated alleles was shown by expression quantitative trait loci analyses in peripheral blood cells. This finding makes *FAM167A* a candidate gene for disease susceptibility. However, the function of FAM167A and the only homologous protein, its gene family member, FAM167B, was unknown when this thesis was initiated.

This thesis aims to elucidate the function of the *FAM167* genes and their role in the pathogenesis of systemic autoimmune diseases. Immunohistochemistry staining of the autoimmune target organ, salivary glands of patients with Sjögren's syndrome, revealed expression of FAM167A in cells both in the inflammatory foci and interstitium. Most were confirmed as B cells, including plasma cells, by double staining. Further, the degree of the FAM167A staining correlated with the focus score, IgG levels, and autoantibodies present in the patients. Investigating the *FAM167* genes and the encoded proteins with bioinformatic methods revealed that they are highly conserved in vertebrates, contain no known protein motifs but a high content of disordered secondary structures. Based on this observation the encoded proteins were denoted disordered autoimmunity (DIORA) -1 and -2. The proteins both localize to the cytosol but are found in distinct immune cell subsets and different organs. To investigate their function at the whole organism level, we established two knock-out mouse strains. Both *Fam167a* and *Fam167b* deficient mice were viable - but *Fam167a* deficient mice showed lower body weight, increased kidney weight, and proteinuria in older animals. Further, alterations in serum immunoglobulins and B cell populations were detected together with an expanded B1a cell population, which exhibited signs of metabolic activation at the transcriptomic level. Contrarily, *Fam167b* deficient mice demonstrated no gross malformations. However, their microglia displayed altered expression of genes belonging to pathways regulated by interferon as well as pathways of chemotaxis, cell adhesion, migration, and inflammatory responses. Based on proximity labeling experiments and annotated protein-protein interactions, we suggest that DIORA-1 and DIORA-2 share MRCKA and MRCKB, both regulators of the cytoskeleton, as interaction partners.

In summary, this thesis presents the initial characterization of two genes with unknown function. Based on the findings, we speculate that the *FAM167* genes contribute to immune function including antibody isotype determination and immunoglobulin production. However, further studies are needed to explore the molecular activity of FAM167A and FAM167B in more detail. In the future, a detailed understanding of the pathways involved in autoimmune pathogenesis such as Sjögren's syndrome, lupus, or rheumatoid arthritis might result in the development of novel therapeutics for the benefit of the patients.

POPULÄRVETENSKAPLIG SAMMANFATTNING

Ett välfungerande immunförsvar skyddar oss från sjukdom, framför allt från infektioner och skadliga ämnen. Om immunförsvaret inte fungerar som det ska kan det bilda antikroppar mot den egna kroppen. Dessa sjukliga processer kan i slutändan leda till kroniska inflammationer och skada frisk vävnad. Till exempel har patienter med Sjögrens syndrom problem med muntorrhet och torra ögon i och med att immunförsvaret förstör saliv- och tårkörtlarna. Utöver körtelproblem kan inre organ och det centrala nervsystemet drabbas av inflammationer. Många patienter med Sjögrens syndrom lider av trötthet, depressioner, led- och muskelvärk samt en förhöjd risk för utveckling av cancerformen lymfom. Generellt begränsas livskvalitén hos patienter med Sjögrens syndrom. De har även ofta en kortare medellivslängd. För närvarande går sjukdomen tyvärr inte att bota. Symptomen går dock att lindra. Därför vill vi med vår forskning uppnå en bättre förståelse av sjukdomen. Genom det hoppas vi att kunna bidra till utvecklingen av nya behandlingsmetoder.

Sjögrens syndrom är en komplex bindvävssjukdom där orsaken till sjukdomsutveckling inte är känd i sin helhet. Vi vet dock att både arvsanlag och miljöfaktorer ökar risken. I studier med många både friska och sjuka deltagare har man kunnat visa på många små variationer i arvsanlagen som ökar risken för sjukdomens utbrott. En av dessa variationer befinner sig mellan arvsanlagen FAM167A och BLK. Denna variation ökar produktionen av FAM167A. FAM167A produceras bland annat av immunceller som har en viktig roll i utvecklingen av sjukdomen. Utifrån detta kan vi dra slutsatsen att arvsanlagen FAM167A predisponerar för Sjögrens syndrom och eventuellt också andra bindvävssjukdomar. När denna avhandling påbörjades var funktionerna för de båda arvsanlagen FAM167A och det besläktade FAM167B helt okända.

Målet med avhandlingen har varit att kartlägga båda arvsanlagens funktioner och deras bidrag till sjukdomsutvecklingen.

Till en början har vi undersökt i vilken utsträckning arvsanlagen förkommer i salivkörtlarna hos patienter med Sjögrens syndrom. Det hittades stora mängder av proteinet FAM167A i immunceller som invaderar salivkörtlarna, inklusive i de immunceller som producerar antikroppar. Dessutom fann vi ett samband mellan mängden FAM167A i salivkörtlarna och antalet invaderande immunceller samt en ökad mängd antikroppar och förekomst av autoantikroppar i blodet.

För att dra slutsatser om betydelsen av de två FAM167 arvsanlagen för en hel organism har vi modifierat arvsanlagen hos möss så att de saknar den ena eller andra av anlaget. Möss utan

FAM167A är livsdugliga men får med tilltagande ålder en lägre vikt och uppvisar dysfunktioner av njurarna. Möss utan FAM167A uppvisar även förändringar vad gäller antikroppsklasser och delpopulationer av immunceller. Hos möss som saknar FAM167B finns generellt färre förändringar. Vi fann dock förändringar av mikroglia, som är en typ av immuncell i hjärnan.

För att förstå den cellulära funktionen av FAM167B har vi identifierat proteiner som interagerar med FAM167B protein i cellen. Bland annat har vi funnit att proteinerna MRCKA och MRCKB är interaktionspartners. Båda dessa protein reglerar cellrörelser.

För första gången presenterar vi här en beskrivning av de två FAM167 arvsanlagens funktion som tidigare har varit okänd. Våra resultat antyder att de har en roll i olika processer såsom cellrörelse och immunfunktion, med särskild betydelse för produktion av antikroppar. Fler studier behövs för att få en mer detaljerad bild av arvsanlagen funktion och deras bidrag till utvecklingen av Sjögrens syndrom. I framtiden kan en bättre förståelse för arvsanlagens funktion möjliggöra utveckling av nya behandlingar för patienter med Sjögrens syndrom och andra inflammatoriska sjukdomar.

POPULÄRWISSENSCHAFTLICHE ZUSAMMENFASSUNG

Ein gut funktionierendes Immunsystem beschützt uns vor Krankheit, insbesondere hilft es bei der Abwehr von Infektionen und anderen Schadstoffen. Funktioniert das Immunsystem aber nicht wie es soll, kann dieses zu der Produktion von Autoantikörpern gegen den eigenen Körper führen. Dieser krankhafte Prozess kann schlussendlich die Zerstörung von gesundem Gewebe und eine chronische Entzündung verursachen. Zum Beispiel leiden Patienten mit Sjögren-Syndrom an Mundtrockenheit und trockenen Augen, da das Immunsystem die Speichel- und Tränendrüsen stark schädigt. Zusätzlich zu der Schädigung der Drüsen kann es zu weiteren entzündlichen Veränderungen in inneren Organen und im zentralen Nervensystem kommen. Viele Patienten mit Sjögren-Syndrom sind zudem von Gelenkschmerzen, Muskelschmerzen, Müdigkeit, Depressionen und einem erhöhten Risiko für die Krebsform Lymphom betroffen. Im Allgemeinen ist die Lebensqualität für die meisten Patienten mit Sjögren-Syndrom stark beeinträchtigt und deren Lebenserwartung häufig verkürzt. Derzeit gibt es leider keine heilende Behandlung für die betroffenen Patienten, stattdessen können lediglich deren Symptome gelindert werden. Deshalb beabsichtigen wir mit unserer Forschung zum grundlegenden Verständnis der Krankheitsentstehung beizutragen, sodass in Zukunft die Entwicklung von neuartigen Therapien möglich ist.

Das Sjögren-Syndrom ist eine komplexe rheumatische Erkrankung und die Ursachen für dessen Entstehung sind im Detail nicht verstanden. Wir wissen jedoch, dass sowohl schädliche Umwelteinflüsse als auch Vererbung zur Krankheitsentstehung beitragen. Studien - an denen sowohl erkrankte als auch gesunde Menschen teilgenommen haben - entdeckten viele kleine Variationen in den Erbanlagen, die ein erhöhtes Risiko für das Entstehen von rheumatischen Erkrankungen verursachen. Eine dieser Variationen liegt zwischen den Erbanlagen FAM167A und BLK. Diese Variation steigert die Produktion von FAM167A. Darüber hinaus wird FAM167A unter anderem in Immunzellen produziert, die eine Hauptrolle bei der Krankheitsentstehung spielen. Auf der Grundlage dieser Befunde schlussfolgern wir, dass die FAM167A Erbanlage zur Prädisposition von rheumatischen Erkrankungen beiträgt. Als diese Doktorarbeit begonnen wurde, waren die Funktionen von FAM167A und deren einzigen verwandten Erbanlage, FAM167B, komplett unbekannt.

Ziel dieser Arbeit ist, die Funktion der beiden FAM167 Erbanlagen und deren Beitrag zur Krankheitsentstehung des Sjögren-Syndroms zu untersuchen.

Als erstes haben wir untersucht, in welchem Umfang das Protein FAM167A in den entzündeten Speicheldrüsen von Patienten mit Sjögren-Syndrom vorkommt. Hohe Mengen des FAM167A Proteins wurden in Immunzellen, die an dem krankhaften Eindringen in die

Speicheldrüsen beteiligt sind und solchen, die Antikörper produzieren, gefunden. Außerdem konnten wir einen Zusammenhang zwischen der FAM167A Menge und der Anzahl der in die Drüse eindringenden Immunzellen sowie mit der Erhöhung von Antikörpern als auch mit dem Vorkommen von Autoantikörpern im Blut herstellen.

Um die Bedeutung der beiden FAM167 Erbanlagen für den ganzen Organismus zu verstehen, haben wir die Erbanlagen von Mäusen so verändert, dass ihnen jeweils eine der beiden fehlt. Mäuse, die kein FAM167A produzieren sind lebensfähig, bekommen jedoch mit zunehmendem Alter ein geringeres Gewicht und zeigen Funktionsminderungen der Nieren. Auch zeigen diese Mäuse abweichende Mengen innerhalb der verschiedenen Antikörperklassen und in einigen Teilpopulationen von Immunzellen. Bei Mäusen ohne die FAM167B Erbanlage sind die Veränderungen generell deutlich weniger, jedoch gibt es Veränderungen in den Mikroglia, einer Sorte von Immunzellen des Gehirns.

Um die Funktion von FAM167B Proteinen innerhalb der Zelle zu verstehen, haben wir interagierende Proteine identifiziert. Unter anderem haben wir MRCKA und MRCKB als Interaktionspartner feststellen können. Diese beiden Proteine sind Regulatoren von Zellbewegungen.

In dieser Arbeit wird zum ersten Mal eine Beschreibung der zwei FAM167 Erbanlagen, deren Funktion unbekannt war, vorgelegt. Unsere Resultate lassen vermuten, dass die FAM167 Erbanlagen eine Rolle in Prozessen wie Zellbewegung und Immunfunktion, insbesondere der Produktion von Antikörpern, spielen. Weitere Untersuchungen sind nötig, um ein detaillierteres Bild von der Funktion beider Erbanlagen und deren Beitrag zur Krankheitsentstehung des Sjögren-Syndroms zu erhalten. In Zukunft wird das bessere Verständnis der für die Krankheitsentstehung verantwortlichen Gene die Entwicklung neuer Therapien für Patienten mit Sjögren-Syndrom und anderen rheumatischen Erkrankungen ermöglichen.

LIST OF SCIENTIFIC PAPERS

I. **Mentlein L**, Thorlacius GE, Meneghel L, Aqrawi LA, Ramirez Sepulveda JI, Grunewald J, Espinosa A, Wahren-Herlenius M. *The rheumatic disease-associated FAM167A-BLK locus encodes DIORA-1, a novel disordered protein expressed highly in bronchial epithelium and alveolar macrophages.* Clin Exp Immunol 2018; **193**:167-77.

II. Aqrawi LA*, **Mentlein L***, Meneghel L, Björk A, Thorlacius GE, Ivanchenko M, Ramírez Sepúlveda JI, Skarstein K, Kvarnström M, Brauner S, Espinosa A, Wahren-Herlenius M. *Clinical associations and expression pattern of the autoimmunity susceptibility factor DIORA-1 in patients with primary Sjögren's syndrome.* Annals of the Rheumatic Diseases 2018. Jul 17. pii: annrheumdis-2018-213634. doi: 10.1136/annrheumdis-2018-213634. [Epub ahead of print]

III. Meneghel L, Ottosson V, Thorlacius GE, **Mentlein L**, Ramírez Sepúlveda JI, Brauner S, Espinosa A, Wahren-Herlenius M. *Generation and characterization of Diora-1 knockout mice.* Manuscript

IV. **Mentlein L**, Meneghel L, Thorlacius GE, Ottosson V, Lund H, Ivanchenko M, Nyberg W, Harris RA, Espinosa A, Wahren-Herlenius M. *Exploring the function of the unknown Diora-2/Fam167b gene in mice.* Manuscript

* Equal contribution

CONTENTS

1 Introduction 1
1.1 A brief overview of the immune system 1
1.2 Autoimmune diseases 2
1.3 Sjögren's syndrome 3
1.4 *The FAM167A-BLK* locus in autoimmune disease 5
1.5 The *FAM167* gene family 6
1.5.1 FAM167A 7
1.5.2 FAM167B 8
2 Aim 11
3 Methodological Considerations 13
4 Results and Discussion 17
4.1 The *BLK-FAM167A* is a susceptibility locus in autoimmune diseases; moreover, eQTL analyses point towards *FAM167A* as a causal gene in disease pathogenesis 17
4.2 DIORA-1 positive cells reside in the foci and interstitial cells within salivary glands of patients with pSS; additionally DIORA-1 staining in the salivary glands correlates with the clinical features focus score, IgG, and Ro/SSA autoantibody levels 19
4.3 The *FAM167* gene family is highly conserved in vertebrates and can be traced to tunicates, indicating a relevant function in multicellular animals 20
4.4 The DIORA protein sequences include no known functional domains and are predicted to contain disordered structures 20
4.5 DIORA-1 and DIORA-2 both localize to the cytosol but have distinct expression patterns within organs and immune cell subsets 21
4.6 *Diora-1* deficient mice present with weight loss, proteinuria, increased IgM levels and altered B cell frequencies, with B1a cells showing metabolic activation 23
4.7 *Diora-2* deficient mice show no obvious phenotypical aberrancies, but their microglia show variable surface marker expression, increased IFN score, and an altered gene signature in cell motility pathways 24
4.8 DIORA-2 acts in proximity to MRCKA, MRCKB, and TXNL1 25
4.9 DIORA-1 and DIORA-2 differ in function but potentially share interaction partners 26
5 Conclusion and Future Perspective 29
6 Acknowledgments 33
7 References 37

LIST OF ABBREVIATIONS

AID	Autoimmune disease
BAL	Broncho-alveolar lavage
BCR	B-cell receptor
BLK	B lymphocyte kinase
CD	Cluster of differentiation
CDC42	Cell division control protein 42 homolog
CDC42BP(A, B)	CDC42 binding protein (A, B)
CNS	Central nervous system
CXCR	C-X-C chemokine receptor
DIORA-1	Disordered autoimmunity 1; synonyms FAM167A, C8orf13
DIORA-2	Disordered autoimmunity 2; synonyms FAM167B, C1orf90
DUF	Domain of unknown function
eQTL	Expression quantitative trait locus
FAM167A	Family with sequence similarity, member A; synonyms DIORA-1, C8orf13
FAM167B	Family with sequence similarity, member B; synonyms DIORA-2, C1orf90
FcγR	Fc gamma receptor
GWAS	Genome-wide association study
HLA	Human leukocyte antigen
IDP	Intrinsically disordered protein
Ig	Immunoglobulin
IL	Interleukin
IFN	Interferon
KO	Knock-out
LCK	Lymphocyte-specific protein tyrosine kinase

MRCK(A, B, G)	Myotonic dystrophy kinase-related CDC42-binding kinase (A, B, G), synonym CDC42BP(A, B, G)
mRNA	Messenger RNA
NFκB	Nuclear factor κB
PBMC	Peripheral blood mononuclear cell
pSS	Primary Sjögren's syndrome
RA	Rheumatoid arthritis
RNA	Ribonucleic acid
ROS	Reactive oxygen species
siRNA	Small interfering RNA
SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
TCR	T-cell receptor
TLR	Toll-like receptor
WASp	Wiskott-Aldrich syndrome protein
WT	Wild-type

1 INTRODUCTION

Autoimmune diseases impose a lifelong burden on the affected patients. Even though our understanding of these complex diseases has expanded, we still do not fully comprehend the disease pathogenesis. Elucidating the molecular pathways behind disease onset and progression will allow us to develop more specific therapies to alleviate the patients' symptoms and to potentially constrain and cure the disease.

1.1 A brief overview of the immune system

The primary task of the immune system is to maintain homeostasis in the organism by clearing pathogens, tumor cells, and dying cells. These functions are executed by a broad spectrum of molecules, cells, and tissues.

The immune system can be divided into two general arms: the innate and acquired/ adaptive immunity. The innate immunity predominately provides a rapid and unspecific defense against pathogens. Contrarily, acquired immunity can adapt to novel threats as it has the potential to memorize, specify and diversify. Thus, it can develop a more efficient immune response than the innate immunity, but the adaptive response usually takes more time to develop. However, the memory allows the adaptive immunity to elicit a specific and more rapid response to pathogens previously encountered.

Granulocytes, monocytes, tissue-resident macrophages, and natural killer cells are the immune cells contributing to innate immunity. B and T cells also have innate functions but are the cells responsible for developing the adaptive immunity. T cells develop in the thymus and are responsible for cell-mediated immunity. B cells develop in the bone marrow and then migrate to lymphoid tissues where they mediate humoral immunity.

In the healthy human, all immune processes are well orchestrated holding infections at bay. Dysregulation within the immune system may result in immunodeficiency, allergies/ hypersensitivity, inflammatory/autoimmune diseases, or cancer.

1.2 Autoimmune diseases

Autoimmune diseases (AID) develop in individuals with uncontrolled immune responses to self-antigens. Between 2 to 5 % of the population in the developed world suffer from this family of disorders [1-3]. A majority of these are rheumatic diseases, in which the chronic inflammatory process leads to tissue destruction resulting in increased morbidity and for most of these conditions, a shorter life expectancy. The onset and progression of autoimmune diseases may be explained by genetic predisposition, environmental factors, and immune regulation [4]. As no curative treatment is available yet, there is a critical need for a deeper understanding of the processes behind chronic autoimmune inflammation.

The current paradigm on AID pathogenesis is that disease may be initiated after environmental triggering in genetically susceptible individuals [5]. The evidence for genetic predisposition stems from studies showing familiar clustering as well as the elevated concordance rate in monozygotic twins [6, 7]. Additionally, many genome-wide association studies (GWAS) have deepened our understanding of the genetic risk contributing to AID. This is discussed in more detail below. Many different environmental triggers have been associated with autoimmunity, in particular, infections, toxic chemicals, dietary components, microbiota and endogenous retro-elements [8-11]. Notably, genes and environment also interact at the level of epigenetics, and several epigenetic modifications have been shown to influence the development and progression of AID [12-14].

How the immune dysregulation causes to break tolerance at the cellular and molecular level may be explained in different ways. B and T cell development is tightly regulated to avoid auto-reactivity, highly self-reactive cells are already deleted centrally, whereas only moderately self-reactive cells are kept in check by immune cells responsible for peripheral tolerance. However, if autoreactive cells escape these checkpoints, they may cause the onset of disease in genetically predisposed individuals [15, 16]. Paradoxically, many autoantigens are intracellular and should thus be hidden or inhibited from binding to antibodies in the extracellular spaces. However, autoantibodies against peptides derived from intracellular proteins are present in many autoimmune patients. Three different processes have been suggested to cause autoantibodies to bind intracellular antigens: apoptosis, antigen modification, and cross-reactivity [17]. More specifically, infections with extensive tissue damage may cause a mechanism called bystander activation where both increased levels of strong pro-inflammatory signals, as well as self-antigens, are present extracellularly, allowing the activation of autoreactive cells [18]. During the process of molecular mimicry,

initially pathogen-derived peptides may resemble self-antigens and activate autoreactive immune cells, which target the host tissue after the original infection is cleared [19]. Additionally, both smoking and bacterial infections have been shown to modify epitopes through post-translational modifications and subsequently may facilitate autoimmunity [20, 21]. Furthermore, the same modification may be recognized on other antigens, resulting in epitope spreading and autoimmunity to more self-antigens [22].

Taken together, aberrant innate, as well as B and T cell responses, together with the production of autoantibodies and pro-inflammatory cytokines, may give rise to autoimmunity. Hence, many immune pathways contribute to autoimmunity including IFN, TLR, and ROS signaling. It may be noteworthy, that several of the pathways involved have also been identified in genetic association studies. The present project was developed to obtain an understanding of the function of a novel gene the research group had identified as associated with Sjögren's syndrome [23].

1.3 Sjögren's syndrome

Primary Sjögren's syndrome (pSS) is a chronic autoimmune rheumatic disease with a prevalence of 0.1 - 0.3 %, affecting women almost twenty times more often than men [24]. The most common symptom is dryness of mouth and eyes, caused by the dysfunction of exocrine glands. Together with manifestations such as fatigue, arthralgia, and increased risk of developing lymphoma, the disease severely impacts the lives of the patients. Anti-Ro/SSA and anti-La/SSB autoantibodies are present in approximately 70 % of patients, and together with histological evidence of inflammation in the salivary glands are used in addition to the clinical symptoms to establish the diagnosis [25].

At the organ level, Sjögren's syndrome is characterized by infiltrates of lymphocytes into exocrine glands, especially into the lacrimal and salivary glands. The infiltration is often accompanied by the production of inflammatory cytokines and autoantibodies by activated B cells [26]. The lymphocytic infiltrates result in reduced or loss of secretory function of the glands, resulting in the cardinal symptoms dry eyes and dry mouth. Beyond this, extra-glandular manifestations such as interstitial pulmonary disease, vasculitis, and thyroid disease are also observed in patients with pSS. Additionally, the incidence rate for lymphomas is increased 15 fold [27]. Serological screenings have revealed that IgM is usually decreased while IgA is increased and IgG highly elevated in patients with pSS. It has

also been shown that the peripheral blood of patients with pSS includes a higher percentage of specific subclasses of activated B cells [28], decreased levels of memory B cells [29], and $CXCR5^+$ cells [30].

Minor salivary gland biopsies are commonly taken during the diagnosis procedure. The focus score is a diagnostic criterion and represents the number of mononuclear cell infiltrates containing at least 50 inflammatory cells in a 4 mm^2 big section of the gland. Several features have been attributed to the B cells in the glandular infiltrates. Ro and La autoantibody-producing cells are found within the salivary gland of patients with pSS [31]. Plasma cells in the foci share characteristics with the long-lived plasma cell subtype [32]. An increase of $CXCR4^+CXCR5^+$ memory B cells within the gland has been confirmed [33].

The etiology of the disease is not fully understood, but as for other rheumatic diseases, it is believed that environmental factors may trigger the disease to develop in genetically susceptible individuals [9]. Based on the different incidence rates in female versus male patients, direct effects due to the sex chromosomes or indirect effects mediated by sex hormones are suspected to play a role in the onset of the disease. One study suggests a role of the X chromosome in the susceptibility to autoimmune disease as men with the Klinefelter syndrome (47, XXY) are more prevalent among male SLE patients [34]. The sex hormone estrogen has been shown to promote an autoimmune phenotype in different mice models [35, 36]. To date, no environmental risk factors have been identified for pSS, but several gene loci have been associated with the disease. A recent GWAS study demonstrated the association to the *HLA*, *IRF5-TNPO3*, *STAT4*, *IL12A*, *FAM167A-BLK*, *DDX6-CXCR5* and *TNIP1* loci at a genome-wide significant level [23]. Previous genetic studies suggested *HLA*, *STAT4*, *IRF5-TNOP3*, *EBF1*, *FAM167A-BLK*, *STAT4* and *TNFSF4* as associated genes [37-41].

The association to the *HLA* genes is the strongest identified so far for all rheumatic diseases. The *HLA* region is one of the most polymorphic regions in the human genome and contains several genes that have been suggested to be involved in disease pathogenesis. The HLA proteins are critical molecules in the adaptive immune response. However, the molecular mechanism behind the association of different *HLA* alleles with a wide range of immunological disorders is not fully understood [42, 43]. In summary, many genetic studies of autoimmune diseases have discovered associations to genes relevant in immune processes.

1.4 *The FAM167A-BLK* locus in autoimmune disease

Among the non-*HLA* loci that have been repeatedly associated with rheumatic diseases is the *FAM167A-BLK* locus. However, the molecular cause behind this association is not understood. Most of the disease-associated SNPs in the *FAM167A-BLK* locus fall into the intergenic region between the two genes, and most of them are positioned closer to the *BLK* gene although the association signal stretches over both genes [23]. Initially, *BLK* has been considered as the cause for the association based on its known role in B cells.

BLK encodes a tyrosine kinase of the Src family, which contributes to B-cell development, differentiation, and signaling [44]. The autoimmunity-associated *BLK* haplotype has been shown to result in altered mRNA and protein expression of BLK in naïve B-cell subsets [45]. Alterations in BCR and NFκB signaling for different SNPs in the *BLK* region have been reported to support the hypothesis that the *BLK* gene is the causal factor behind this association [46]. However, rather surprisingly, *Blk* knock-out mice show no overt immune phenotype [47]. Only in $Blk^{-/-}Fyn^{-/-}Lyn^{-/-}$ triple knockout mice an altered B cell development has been observed, indicating that these Src kinases can compensate each other's function [48]. On the other hand, *FAM167A* (family with sequence similarity 167, member A) is expressed in a broad range of cells and tissues according to the human protein atlas (https://www.proteinatlas.org), but its function remains elusive.

Moreover, the genetic association to the *FAM167A-BLK* locus has not only been published for pSS [23, 38] but has been shown for several autoimmune diseases (Figure 1). For systemic lupus erythematosus (SLE), Hom and colleagues [49], were the first to report the association at the genome-wide level, followed by others [50-54]. The association to the *BLK-FAM167* locus has also been shown in several studies involving rheumatoid arthritis (RA) patients [55-57]. Kawasaki disease is an acute vasculitis primarily affecting children. The strongest association found in a GWAS of Kawasaki disease by Onouchi et al. [58] is the rs2254546 SNP in the *FAM167A-BLK* region. This association has also been confirmed below genome-wide significance level in other studies [59, 60]. Interestingly, the signal from the *FAM167A-BLK* locus is higher than that from the *HLA* locus in Kawasaki disease. For other traits, the association with SNPs in the *FAM167A-BLK* locus has been shown below genome-wide significant level for instance in polymyositis/dermatomyositis [61], systemic sclerosis [46, 62, 63], antiphospholipid syndrome [64], and IL-6 cytokine production in response to vaccination [65]. However, for giant cell arteritis, an association could not be confirmed in a study genotyping two of the representative SNPs in the region, but one allele

of a risk SNP could be correlated to an increased incidence of severe ischemic complication [66]. Besides the association with autoimmune and rheumatic diseases, two case-control studies in Chinese patients with asthma or allergic rhinitis show an association to the *FAM167A-BLK* locus [67, 68]. Further, a bigger genomic fragment, which includes among others also the *FAM167A-BLK* locus, has been associated with keratolytic winter erythema [69]. Overall, most of the disease-associated SNPs fall into a small distinct area around SNP rs2736240. More specifically, a meta-analysis of that specific SNP rs2736240 also concludes that this locus associates to various autoimmune diseases [70]. In most of the studies of pSS, SLE and RA the odds ratio for the *FAM167A-BLK* locus was determined around 1.3. In the Kawasaki study by Onouchi et al the mean of the odds ratio is somewhat higher for this condition at 1.8 [58].

Further, the 8p23.1 deletion syndrome that compromises a loss of a 2.4 Mb region including several genes in addition to *FAM167A* is associated with conotruncal heart defects, microcephaly and global developmental delay [71]. 8p23.1 also represents one of the most extensive known inversions, a common polymorphism, and the association to the risk allele in the *FAM167A-BLK* locus persists independently of the inversion haplotype in SLE patients, however, the non-inversion haplotype is associated with SLE and might additively affect the pathogenesis of SLE [72].

Interestingly, *cis*-eQTL analysis shows a highly increased expression of *FAM167A* for the genotype associated with pSS [23]. Notably, the effect on *BLK* expression is moderate/low. Similar effects were also shown for SLE-associated polymorphisms were the disease-associated genotype resulted in substantially elevated mRNA expression of *FAM167A* and somewhat increased expression of *BLK* in transformed cell lines [49]. In line with these findings, a recent study found a powerful eQTL effect on FAM167A expression in B cells of RA patients [73].

1.5 The *FAM167* gene family

The genes of the family with sequence similarity 167 (*FAM167*) are grouped by their sequence resemblance. We found that the *FAM167* gene family consists of two members, *FAM167A* and *FAM167B*. Neither of them shows homology to any other annotated genes nor encodes any previously known protein domains or motifs. *FAM167A* is located on chromosome 8p23.1. Based on its genetic position the *FAM167A* gene is also denoted as

C8orf13. The gene compromises four exons and encodes a protein of 214 amino acids. The *FAM167B* gene is located on chromosome 1p35.1 and therefore also annotated as *C1orf90*. *FAM167B* contains two exons encoding for a smaller protein of 163 amino acids. Initially, the *FAM167* genes were denoted as *SEC* based on their partial similarity to the sequence of an oncogene from a breast cancer cell line published 1990 [74]. However, no such gene is currently annotated in the NCBI database. *FAM167A* was mapped and sequenced for the first time in a study on the keratolytic winter erythema locus [69], which also included the previously cloned human *BLK* [75].

1.5.1 FAM167A

When this thesis was initiated, there were no publications directly addressing *FAM167A* expression, localization, or function. However, there were several reports on genetic disease associations within the region.

FAM167A is transcribed in many cell lines and tissues. Interestingly, the SNPs within the *FAM167A-BLK* region that are associated with pSS and SLE locate to the same region targeted by epigenetic modifications in B cells, indicating that expression differences may indeed relate to immune cells and the disease [76, 77]. *FAM167A* is also one of several genes that are hypo-methylated and upregulated in pilocytic astrocytomas compared to diffuse astrocytomas [78]. Another association with cancer comes from *FAM167A* being positioned within one of the 47 putative long-range epigenetic silencing regions in prostate cancer [79]. CNV (copy number variation) was also found within *FAM167A-AS1 (C8orf12)* in patients with disorders of sexual development [80]. *FAM167A* is also one of the early genes expressed in the ventricular zone in brains of ferrets [81], and in a co-culture of bone marrow stromal cells of healthy donors with the myeloid leukemia cell line K562, *FAM167A* was in the top ten of upregulated genes [82]. Furthermore, *FAM167A* was one of four genes to be associated with telomerase length throughout aging [83]. All these findings are rather vague, therefore only a more targeted approach will help to elucidate the function of *FAM167A*.

1.5.2 FAM167B

FAM167B is the other member of this gene family. Its functional role was not understood when this project was launched. High transcription of *FAM167B* has been reported in publically available databases in melanoma cell lines (cancer cell line encyclopedia, https://portals.broadinstitute.org/ccle) and for adrenal, kidney and thyroid tissue (GTEx, https://gtexportal.org/). Generally, the *FAM167B* gene and protein are scarcely covered in the scientific literature. Interestingly, one study characterizing the pre-integration complexes of HIV (human immunodeficiency virus) within infected cells found FAM167B is one of the top ten proteins unique to the infected samples [84]. Thus, one could speculate that FAM167B is a host factor directly or indirectly involved in HIV replication. Additionally, *FAM167B* was also found to be one of 72 host factors required for VSV infection through a big siRNA screen [85]. Going hand in hand with the finding of high *FAM167B* transcription in melanoma cell lines, two independent studies show increased *FAM167B* transcription in melanoma and both studies are indicating an association to more aggressive tumors [86, 87]. *Fam167b* was observed to be one of 50 up- and downregulated genes in pancreatic islets of *Ffar3*$^{-/-}$ mice that lack the free fatty acid receptor-3 gene [88]. *FAM167B* has also been found as one of many genes upregulated upon IL-6 and IL-6 trans-signaling in human airway smooth muscle cells [89]. *Fam167b* was one of the genes characterized by the ingenuity pathway analysis associating a network of genes involved in inflammatory processes with the overexpression of miR-30b in the mouse mammary gland [90]. *Fam167b* is also one of 1100 genes differently regulated in bone marrow-derived small embryonic-like pluripotent stem cells exposed to intermittent hypoxia [91]. *FAM167B* potentially, together with its neighboring genes, forms a risk locus for the eating disorder anorexia nervosa ($p=9.6 \times 10^{-6}$) [92]. All these findings are rather indicative but might be used to from working hypotheses while exploring the biological function of *FAM167B*.

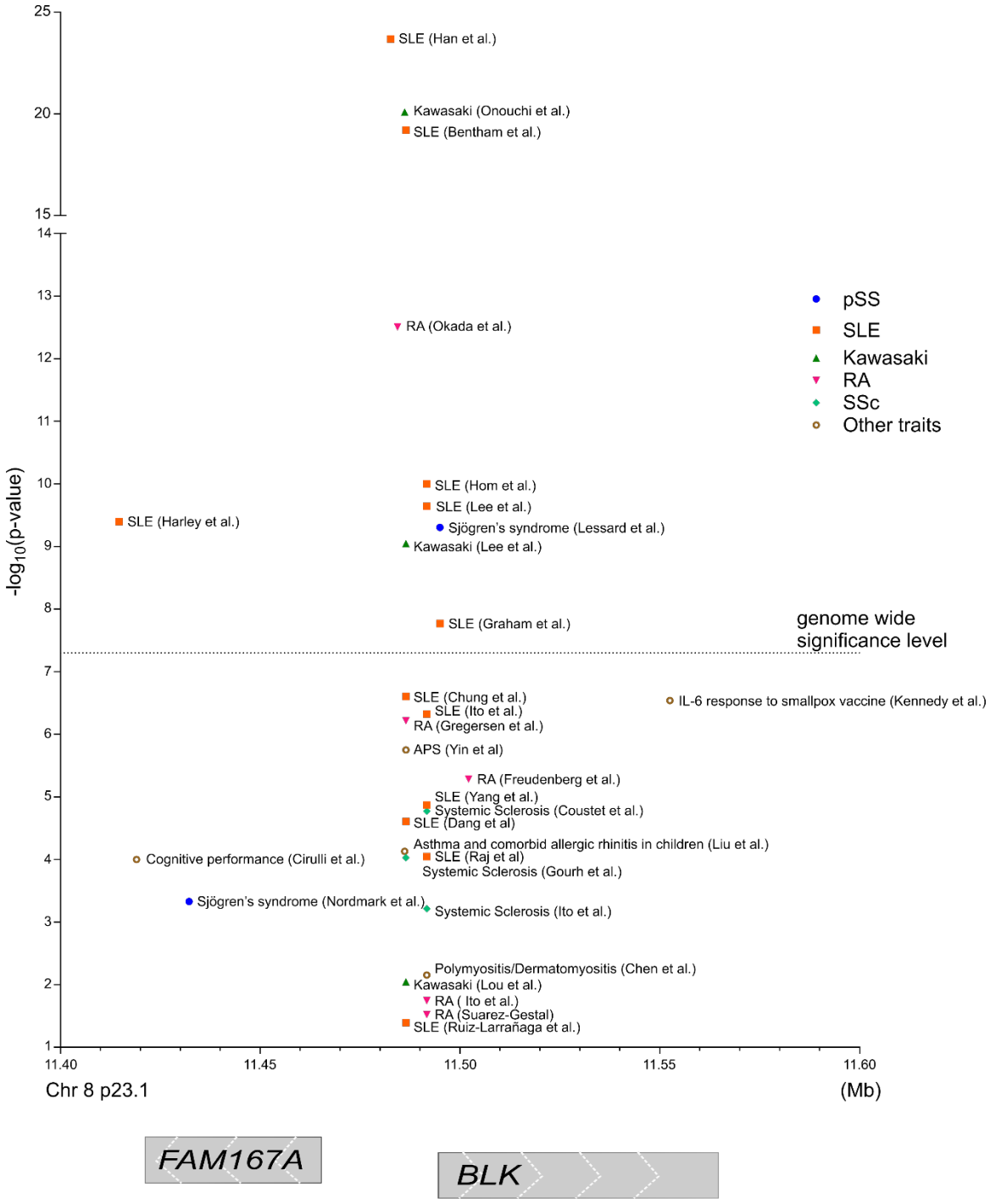

Figure 1: Genetic studies showing an association to the *FAM167-BLK* locus are denoted with the position of the most significant SNP and their respective p-value, the label further includes the relevant disease or trait and the first author of the study.

2 AIM

Genetic association studies in pSS, SLE, RA, and Kawasaki disease have demonstrated that individuals with specific polymorphisms in the *FAM167A-BLK* locus are more prone to develop these diseases. eQTL analyses have shown moderate effects of associated polymorphisms concerning the expression of *BLK*, while the genotype has a significant impact on the expression of *FAM167A*. *BLK* (B lymphocyte kinase) has been identified as acting downstream of the B cell receptor, while no function has yet been assigned to *FAM167A*. This thesis project aims to characterize the *FAM167* gene family and to gain a detailed understanding of the two gene family members, *FAM167A* and *FAM167B*. Specifically, the project entails:

- Gathering and interpreting information on the *FAM167* family genes and translated products from public databases to understand their evolution, conservation, expression, and protein features (Paper I)

- Defining organ and cellular expression of FAM167A and FAM167B in different organs and cells (human and mouse) (Paper I and II)

- Test for FAM167A expression in autoimmune target organs, understand staining pattern and correlate with clinical parameters of the patients (Paper II)

- Generation of *Fam167a* and *Fam167b* knock-out mice to understand the importance and function of the genes (Paper III and IV)

- Identify interaction partners of the FAM167B protein to further understand its role within the cell (Paper IV)

3 METHODOLOGICAL CONSIDERATIONS

The experimental methods which are central to the findings presented in this thesis are discussed with a focus on advantages and limitations in this section. The detailed description of the methods used is available in the respective papers attached.

Localization study with transfected YFP-fusion protein

Localization experiments were performed by tagging our proteins of interest with another fluorescent protein. The fluorescent fusion protein can be easily detected in the microscope and marks the localization of the protein of interest. The main advantage of this approach is that it only requires the cloning of the fusion protein, all other reagents are commercially available. Furthermore, the expression construct is delivered by transfection, which can be directly implemented in a wide range of cell lines. Unfortunately, the transfection usually results in high expression and may induce artifacts or aggregates. Besides, every genetic modification, here the tagging, may cause alterations to the protein, both in function or localization as the tag may conceal or block relevant protein structures. In sum, this kind of localization study is easily implemented but should be complemented by an alternative approach to prove the biological relevance of the findings.

Immunohistochemistry

This method detects the distribution of specific antigens in tissue sections or cell suspensions on microscope slides. The detection is mediated by antibodies specific for the target of interest. Next, the staining can be visualized by established secondary antibodies detecting the constant domain and which are coupled to a fluorescent dye or an enzyme catalyzing a colorimetric reaction. Therefore, this method is strictly dependent on the availability of specific antibodies. Often the use of monoclonal antibodies is preferred as they recognize only one epitope. Different experimental conditions may change the conformation of the epitope, and thus the employability of the monoclonal has to be confirmed in each experimental set-up.

Currently, there are no monoclonal anti-FAM167A or FAM167B antibodies available. However, the human protein atlas consortium has developed several affinity-purified

polyclonal antibodies. We tested all commercially available antibodies on cell lysates containing overexpressed tagged FAM167 proteins. We found that the HPA030426 anti-FAM167A antibody exhibited the smallest amount of non specific binding and reliably stained its target. Its cross-reactivity to the mouse Fam167a protein was not as strong as to the human protein, and therefore we restricted our investigations of FAM167A by immunohistochemistry to human tissues and cells. A disadvantage of using polyclonal antibodies for immunochemistry may be the lesser specificity resulting in background staining, but they may be less sensitive to conformational changes of the epitope as they recognize many epitopes. Luckily, our antibody of choice exhibited no detectable background staining and enabled us to identify a FAM167A staining pattern discriminating samples from patients with pSS versus non-SS. To conclude, immunohistochemistry allows localization and expression studies in a broad variety of cells and tissues but is critically dependent on the availability and quality of the employed antibodies.

Fam167a knock-out mouse

To elucidate the unknown function of *Fam167a*, we generated a knock-out (KO) mouse strain. Based on the association to AID and expression in various immune cells we suspected a function of *Diora-1* in the immune system. The human *FAM167A* shares high sequence similarity with the murine *Fam167a* as expected for paralogs. The murine immune system shares the main features with the human immune system. Therefore, mice became central to immunological research: many relevant disease models and tools have been established. Furthermore, we chose the classical approach of deleting the gene of interest to understand which processes are disturbed in its absence.

To generate *Fam167a* deficient mice we employed the by the International Mouse Phenotyping Consortium [93] predesigned construct Fam167a$^{tm1(KOMP)Vlcg}$. In this construct, the gene is entirely deleted. An inserted LacZ reporter allows longitudinal expression profiling of the gene independently from the availability of specific antibodies.

However, the genomic deletion may alter relevant regulatory sequences or disturb other coding sequences in or nearby the modified locus. Due to the complexity of the genome, it may be complicated to prove that the observed phenotype is not caused by the deletion of other relevant genomic regions. Nevertheless, cellular knock-down experiments which target genes post-transcriptionally could be complementary to confirm the function of the gene.

Fam167b gene trap mouse

For the generation of *FAM167b* deficient mice, we used the Fam167b$^{tm1a(EUCOMM)Wtsi}$ construct also developed by the International Mouse Phenotyping Consortium [93]. This design is called 'KO first allele (reporter-tagged insertion with conditional potential)'. In detail, the insertion of additional sequences disrupts the splicing of the two exons and therefore the correct expression of the gene. Exon 1 is instead spliced to the LacZ reporter which allows expression profiling. This concept is called gene trap and is assumed to abolish the gene expression similarly to classical KO mice. However, a leaky expression is theoretically possible and it is advised to confirm the absence of gene expression and its product. In our mouse strain, we confirmed the lack of *Fam167b* transcripts as no anti-mouse Fam167b antibody is available.

Additional inserted genomic sequences allow that the gene trap may be reverted through breeding with the correct Flip mouse strain. In an additional breeding step with a specific Cre mouse, one may obtain mice with deficient expression in selected tissues or cell populations. Hence, the name of the construct, first, it resembles a KO but may be developed to a conditional mouse strain.

Proximity labeling assay (BioID2)

A protein method that uncovers proteins in the vicinity of the protein of interest is highly relevant as this allows the description of protein complexes. This method has been refined recently by the development of a promiscuous biotin ligase of reduced size, named BioID2 [94, 95]. The ligation of the small biotin molecule to proximal proteins allows the labeling of weak or only transient interactions and subsequently these labeled potential interaction partners may be easily extracted.

As discussed before, the genetic modification necessary to generate the fusion protein consisting of the BioID2 ligase and the protein of interest may alter the function or localization of the protein. Further, the tagged protein may induce steric hindrance. To overcome these limitations, alternative cloning strategies may be tested, e.g. adding the tag to the other end of the protein, shorten or prolong the linker between tag and protein of interest.

In our experiments, we chose a cell line that endogenously expresses FAM167B assuming that its natural interaction partners should be expressed in the same cell also. To avoid artifacts caused by overexpression, we confirmed that the transduction resulted in very low expression of the fusion protein. The utilized biotin ligase has slow kinetics. This may result in the

labeling of proteins responsible for its biogenesis, transport, or other processes, which are not necessarily relevant for its function. Therefore, conducting several experiments or such experiments with the potential for quantification of the hits may promote the discovery of relevant hits.

This method has the potential of uncovering protein complexes and therefore elucidating functional networks. Ideally, those candidate novel interactions discovered will be confirmed by other methods like co-immunoprecipitation, pull-down, or crosslinking. After that other molecular protein or protein structure tools can be utilized like targeted mutagenesis to define interaction domains, surface plasmon resonance to measure the quality of the interaction, and structure determination of the interaction complex.

4 RESULTS AND DISCUSSION

4.1 The *BLK-FAM167A* is a susceptibility locus in autoimmune diseases; moreover, eQTL analyses point towards *FAM167A* as a causal gene in disease pathogenesis

The rheumatic diseases systemic SLE, pSS, and RA are associated with polymorphisms in the *FAM167A-BLK* locus (Paper I, Figure 1). In large GWAS studies of SLE, RA and pSS this association is above genome-wide significance and the increased risk results in odds ratios between 1.2 - 1.3 [23, 49, 55]. These risk SNPs are in high linkage disequilibrium. Therefore, we assume these represent the same genetic signal. In smaller studies of polymyositis/dermatomyositis [61], systemic sclerosis [46, 62, 63], antiphospholipid syndrome [64], IL-6 cytokine production after vaccination [65] and allergy [67, 68] associations to the *FAM167A-BLK* locus could also be detected, though, at lower significance levels. Interestingly, in Kawasaki disease, the genetic association with the *FAM167A-BLK* locus is stronger than with the *HLA* locus reflected in a high odds ratio of 1.8 [58]. Additional studies will be necessary to understand how *FAM167A* influences the inflammation of blood vessels in the affected patients. While a wide array of genetic studies show associations to the *FAM167A-BLK* locus, the causal SNP has not been identified yet. Overall, the *FAM167A-BLK* locus is associated with several immune disorders with substantial evidence to be a susceptibility locus in autoimmune and inflammatory diseases.

Further, we show a strong eQTL effect of the disease-associated polymorphism on the transcription of *FAM167A* in PBMC (Paper I, Figure 1). This effect is even stronger when investigating purified B cells [96]. However, we could not detect that the same polymorphism affects *FAM167A* transcription in monocytes. In line with our findings, the previous GWAS of pSS also found the same eQTL effect in PBMC [23]. Before, the first SLE GWAS investigating an eQTL effect at the *FAM167A-BLK* locus found an increment in *BLK* transcription, but an even more prominently increased *FAM167A* transcription related to autoimmunity-associated alleles [49]. A recently published analysis of 101 established risk loci in RA in both T and B cells determined the strongest eQTL effect in B cells on the transcription of *FAM167A* [73]. This eQTL study also shows an effect on *BLK* transcription in RA patients. However, this effect is rather moderate compared to the significance of allele-related *FAM167A* transcription levels in B cells, although the differential allele-specific BLK transcription in T cells was highly significant. In summary, we conclude that several studies

show a significantly increased *FAM167A* transcription by the disease-associated alleles of specific polymorphism in several rheumatic diseases.

The disease-associated polymorphisms fall into the intergenic region between *FAM167A* and *BLK*. As BLK is expressed in B cells, acting downstream of the BCR, it was long considered the culprit behind the association. We found a significant eQTL effect on the transcription of *BLK* with lower expression from disease-associated alleles; however, the effect size on *FAM167A* transcription is greater. These findings do not exclude that *BLK* contributes to disease pathogenesis, but based on the increased transcription of *FAM167A* related to the disease-associated alleles we decided that this uncharacterized gene with unknown function deserves a more profound analysis to understand its potential contribution to disease pathogenesis.

One, at first sight, conflicting result is that we do not see increased transcription of *FAM167A* in the PBMC of our patient cohort compared to healthy donors (Paper II, Figure 1). However, the analysis is not stratified according to genotype, and as for many autoimmune-associated genotypes, it is common also among the healthy population. Thus, the effect of the risk SNP on *FAM167A* transcription might be diluted in this small cohort of mixed genotypes at the risk locus.

Furthermore, we detected a variable transcription of *FAM167A* in B cells of healthy donors, dividing them into two almost equally big groups with either low or elevated *FAM167A* transcription. One could speculate that genetic variability within these donors may contribute to the variable transcription. In addition to the genotype, also environmental factors or the current immune status of the donors may contribute to differences in the *FAM167A* transcription. Therefore, it would be interesting to understand whether *FAM167A* is a marker, being regulated or a regulator of immune function in B cells.

Besides, it is highly relevant, that the variable *FAM167A* exp transcription ression is observed in B cells, which are important to autoimmunity. These initial data that not only identified a genetic association with *FAM167A* but also intrinsic and potentially extrinsic regulatory effects on transcription of *FAM167A* make it an interesting candidate gene to contribute to disease pathogenesis.

4.2 DIORA-1 positive cells reside in the foci and interstitial cells within salivary glands of patients with pSS; additionally DIORA-1 staining in the salivary glands correlates with the clinical features focus score, IgG, and Ro/SSA autoantibody levels

To increase our insight into the role of *FAM167A* and its gene product, the DIORA-1 protein, in autoimmune disease pathogenesis we stained salivary gland biopsies of patients with Sjögren's syndrome for DIORA-1. DIORA-1 positive cells were found both in lymphocytic foci and interstitially in the glands. Moreover, the ductal epithelium stained also positively for DIORA-1. The same staining of non-SS patients' salivary glands hardly identifies any cells expressing DIORA-1. Further, a correlation between the staining and focus score was observed (Paper II, Figure 1). Based on this finding we assume that mainly the infiltrating immune cells stain positively for DIORA-1.

To further characterize the DIORA-1 positive cells we performed double staining with CD20, CD38, and CD138. A majority of the DIORA-1 positive cells were determined as B cells. Also, a fraction of the DIORA-1 positive cells was identified as plasma cells. This finding is coherent with the transcription of *DIORA-1* in B and plasma cells in circulation which was confirmed by qPCR. However, plasma cells in circulation are not high expressers of *DIORA-1* whereas DIORA-1 staining is intense in the plasma cells of the inflamed glands of patients with pSS. Considering their distinct localization, these plasma cells are not per se comparable as they represent two distinct plasma cell subsets and different levels of DIORA-1 expression might depend on their further differentiation in the tissue.

Moreover, the amount of the DIORA-1 staining correlates with the parameters hypergammaglobinemia and the presence of Ro/SSA autoantibody in the serum. Based on this connection, we speculate that DIORA-1 has a direct impact on antibody levels potentially by augmented production or promotion of the IgG class switch. Plasma cells in target organs of autoimmunity have been identified to produce autoantibodies [31, 97]. Therefore, future experiments might test for the antigen specificity of DIORA-1 positive plasma cells to investigate if DIORA-1 expression correlates to the production of autoantibodies.

In our initial study, DIORA-1 staining did not correlate with patient gender, age, salivary flow rate, extraglandular manifestations, or adverse predictors of lymphoma (salivary gland enlargement, rheumatoid factor, and lymphopenia). However, to address the potential of DIORA-1 staining as a possible novel biomarker or prognostic factor, a larger longitudinal study would need to access this in more depth.

In summary, these findings suggest that DIORA-1 operates in different B cell subsets and give additional indirect evidence for the contribution of DIORA-1 to the inflammatory process.

4.3 The *FAM167* gene family is highly conserved in vertebrates and can be traced to tunicates, indicating a relevant function in multicellular animals

We found that homologs to both *FAM167A* and *FAM167B* are highly conserved (Paper 1, Figure 2). This circumstance does not only allow us to study these genes in classical model organisms like mice or zebrafish but also indicates that these genes have a relevant function considering that they have not been erased from vertebrate genomes throughout the course of millions of years.

The *FAM167* gene family formed before vertebrates developed features of adaptive immunity. However, we observed expression in B cells and alterations in the adaptive immune system of *Fam167a* deficient mice. Therefore, we speculate that the gene may have either more than one function or an essential function within the cell also affecting pathways that evolved later.

In the evolutionary tree, tunicates represent animals with a primary innate immune system consisting of features like pattern recognition receptors and phagocytic cells. Thus, we additionally hypothesize that the genes could have co-evolved with the innate immune system, contributing to essential functions within the more complex immune system of vertebrates.

Furthermore, we found that several neighboring genes of both *FAM167* genes share high homology, suggesting that this genetic fragment was duplicated early during evolution. Gene duplication is widespread in more complex genomes and seminal for the diversification of function. Hence, the duplication may suggest that the genes within the two loci may have acquired different functions through evolution and/or are regulated differently.

4.4 The DIORA protein sequences include no known functional domains and are predicted to contain disordered structures

FAM167A encodes for a protein of 24 kDa and FAM167B for a smaller protein of 18 kDa. At the protein sequence level, the domain of unknown function (DUF3259) is annotated at the C-terminus of both proteins. This shared domain is the reason for grouping the two genes into one family, although high sequence similarity can be observed also at the more N-terminal parts of the proteins. No molecular function has been linked to the DUF3259

sequence motif. Besides, little is known on the proteins. The sequence is predicted to contain no transmembrane or localization sequences. In consequence, they are predicted to localize to the cytosol.

The homology analysis of FAM167 proteins from thoroughly sequenced model organisms detected conserved sequence stretches both at the N- and C-terminus with a less conserved linker sequence in between (Paper I, Figure 3). Secondary structure predictions reveal a large degree of disorder, mainly in the N-terminal part of both proteins. Overall, the pattern of the disorder is similar in both proteins (Paper I, Figure 4). However, the C-terminus comprises more helical and coil-coiled sequence structures.

Based on the evidence of an association with autoimmunity and the disordered structure we denoted this protein/gene family as “Disordered Autoimmunity” DIORA. Proteins with disordered secondary/tertiary structures are loosely grouped as “Intrinsically Disordered Proteins” (IDP). A large fraction of eukaryotic proteins contains disordered structural elements and especially yet uncharacterized proteins are more frequently disordered compared to those already annotated, making it a class of proteins with a high potential for discovering novel functions [98]. Many IDPs contribute to intracellular signaling and regulatory processes, and many of them function as central interaction hubs within cells. Gsponer and Babu classify IDPs according to their functional and structural characteristics such as “(i) their facilitated regulation via diverse post-translational modifications of specific amino acids (ii) scaffolding and recruitment of different binding partners in space and time via the "fly-casting" mechanism, through peptide motifs and by coupling folding with binding and (iii) conformational variability and adaptability” [99]. As the DIORA proteins contain vast stretches of flexible structure, they possibly execute their function via interaction with several proteins and consequently regulate cellular pathways.

4.5 DIORA-1 and DIORA-2 both localize to the cytosol but have distinct expression patterns within organs and immune cell subsets

Next, we investigated the expression pattern of the DIORA proteins. By understanding where a protein is present, we obtain an initial understanding of how it operates. Overexpression of both DIORA proteins shows a dotted pattern within the cell cytoplasm. Overexpression of proteins intracellularly may lead to aggregates and artefactual patterns, but a similar pattern was confirmed for DIORA-1 by antibody staining of non-transfected primary bronchoalveolar lavage (BAL) cells (Paper I, Figure 5). The pattern resembles the distribution

of endosomes; confirmation of the co-localization would, however, require co-staining with adequate markers.

Further, more detailed localization studies on DIORA-2 will require a highly specific antibody to be produced. Currently, the human protein atlas annotates DIORA-2 to co-localize with the actin skeleton. This localization may be explained by our finding that two interactors identified below are regulators of the cytoskeleton. However, this is in conflict with our overexpression data, but the intracellular pattern might also vary in different cell types or fluctuate depending on external signals, activation station, cell division, or other factors. The production of specific monoclonal antibodies will be required to refine the intracellular expression of the DIORA proteins. In sum, our initial data suggest that both DIORA proteins are expressed within the cytosol.

The highest transcription of *Fam167a* was observed in the lungs of mice, but it was also expressed in many other organs including spleen, skeletal muscle, brain, liver, thymus, lymph node, and heart. The high expression in the lung was confirmed in human lung biopsy and BAL cells (Paper I, Figure 5). The staining of the reporter gene in the lungs of *Fam167a* deficient mice confirms an epithelial pattern (Paper III, Figure 1). According to the ImmGen database (Figure 2, www.immgen.org), *Fam167a* is transcribed in many immune cells, including B and B1a cells. According to the cancer cell line encyclopedia (https://portals.broadinstitute.org/ccle), FAM167A is transcribed in a wide variety of cell lines and unusually high in the myeloma cell line LP-1 which we also observed (Paper II, Figure 1C).

The lung has been suggested to contribute to the development of AID as exposure to smoke or other airborne triggers correlates to an increased incidence rate in RA and SLE. These matters are suspected to repetitively activate the immune system through substances in the smoke or the increased infection rate and subsequently facilitate the onset of AID. Therefore, it would be interesting to explore further the function of FAM167A expression in the lung epithelium and immune cells in order to investigate its potential role in triggering of autoimmunity.

The transcription profiles of *Fam167a* and *Fam167b* in mice scarcely overlap. The highest transcription of Fam167b was detected in the adrenal gland, moderate transcription was observed in kidney and liver. Both the ImmGen (Figure 2) and other databases show specifically high transcription of Fam167b in microglial cells. According to the cancer cell line encyclopedia FAM167B has specifically high transcription in human melanoma cell lines. To

conclude, both genes show distinct expression profiles at the organ level, and both are transcribed in distinct immune cell subsets.

4.6 *Diora-1* deficient mice present with weight loss, proteinuria, increased IgM levels and altered B cell frequencies, with B1a cells showing metabolic activation

For investigating the function of *Diora-1* within the whole organism, we established a knock-out mouse (Paper III). *Diora-1* deficient mice were viable, and no embryonic lethality or shortened life-expectancy was observed. Further, no major malformations or pathohistological abnormalities were discovered. In old (26 weeks or more) male *Diora-1* deficient mice, we noted decreased body weight, increased kidney size, and elevated protein levels in the urine. Also, no differences in peripheral blood counts could be detected. However, immunoglobulin levels were altered in serum. While IgA was decreased, elevated levels of IgM were measured, and higher numbers of splenic IgM producing cells were observed. IgG levels did not differ between WT and knock-out mice. Characterizing immune subsets in the spleen revealed increased levels of B cells, in particular, B1a, in *Diora-1* deficient mice compared to wild-type mice. Similarly, IgM levels were elevated *Diora-1* deficient mice upon vaccination. Transcriptomic profiling of splenic B1a cells detects significant alterations in the gene signatures of oxidative phosphorylation, fatty acid metabolism, and *mTorc1*.

One explanation for the decreased body weight could be that IgA deficiency in the gut has been linked to defects in the lipid metabolism [100]. Alternatively, the weight loss might be explained by the malfunctioning of the kidney, the leakage of protein into the urine or the ongoing inflammation in these mice.

To understand the altered B cell frequencies in *Diora-1* deficient mice, it would be highly relevant to investigate if extrinsic or intrinsic factors cause this dysregulation. Therefore, one could employ lineage-specific knock-out models. Considering that *Diora-1* deficient mice produce elevated IgM antibodies but reduced levels of those more matured immunoglobulins, the first thing that comes to mind is a malfunction in the class-switch process. However, as normal IgG levels were observed, this would be partial or specific for class-switch related to generating IgA. An alternative explanation is that cells preferentially producing IgM are expanded in *Diora-1* deficient mice. This may be explained by the increased B1a cell population in *Diora-1* deficient mice. Moreover, future experiments may test, if the IgM molecules of *Diora-1* deficient mice are autoreactive.

The transcriptomic profile we see in the B1a cells of *Diora-1* deficient mice resembles the gene signature in metabolically activated cells. Recently, the field of immune metabolism has received increased attention, with the understanding that activated T and B cells shift to the more energy costly but faster metabolic process of glycolysis. In a lupus mouse model, the disease could be reversed by antagonizing glycolysis in $CD4^+$ T cells, proving the relevance of immune metabolomics for developing novel therapeutic strategies for AID [101].

Taken together, we see signs of inflammation or immune dysregulation in *Diora-1* deficient mice. In the *Diora-1* knock-out mice, we observed an increase in IgM and decrease of IgA. The serological pattern is also disturbed in patients with pSS compared to healthy donors. Specifically, they present with decreased IgM, increased IgA and very high IgG levels, especially of the IgG1 subtype [102]. Therefore, it might be relevant to test further if the IgG subclasses are disturbed in *Diora-1* deficient mice.

However, one should keep in mind that the disease-associated SNP results in increased *DIORA-1* transcription. Thus, the increased *DIORA-1* expression is assumed to contribute to disease pathogenesis. A *Diora-1* overexpression mouse model could be implemented, eventually in combination with a disease model, to investigate if the gene explains some of the features typical for the human disease.

4.7 *Diora-2* deficient mice show no obvious phenotypical aberrancies, but their microglia show variable surface marker expression, increased IFN score, and an altered gene signature in cell motility pathways

Diora-2 deficient mice were generated, and the functionality of the gene trap confirmed (Paper IV). The mice were viable and showed no gross malformations. We conclude, therefore, that the gene is not vital in development, limiting the cells expressing it or that the mice may compensate for its function through other regulatory mechanisms. Further, no differences in peripheral blood counts were detected, which goes in hand with low/non detected expression in the cell types assessed. However, the investigation of a panel of 92 serum proteins found significant variations between wild-type and *Diora-2* deficient mice. In young mice, we see upregulation of S100a4 and IL-10. High levels of these proteins in serum have been previously associated with AID [103, 104]. The other serum proteins, which are significantly upregulated in *Diora-2* deficient mice, function in a wide range of processes including immune regulation and cell migration. In old *Diora-2* deficient mice, 15 out of 92

serum proteins are significantly downregulated. This may reflect a secondary effect resulting in the dysregulation of several pathways due to the prolonged absence of Diora-2.

Next, we characterized the myeloid immune cell populations in the brain, as we had learned from public databases that microglia are specifically high in *Diora-2* transcription. We observed no changes in the cell counts, but the surface markers CD11c (ITGAX), CD11b (ITGAM), and CD45 (PTPRC) were elevated on the microglia of *Diora-2* deficient mice. Taking into account that the logic for investigating DIORA-2 results from the genetic association of its only other gene family member DIORA-1 with rheumatic autoimmune diseases, it may be noteworthy that both ITGAM and ITGAX are genetically associated with SLE [49]. Furthermore, the risk allele of CD11b (ITGAM) correlates with increased INF-I levels in SLE patients [105]. Considering that Diora-2 localizes to the cytosol, we assumed that the increase of the surface marker expression is a secondary result of a dysregulation within the microglia of *Diora-2* deficient mice. Therefore, we performed a microarray analysis to understand the transcriptional regulation in more detail. We detected many significantly differentially expressed genes in the microglia of *Diora-2* deficient mice. In particular, the gene enrichment analysis points to alterations in the pathways of positive regulation of immune responses, leukocyte migration, cell-cell adhesion, granulocyte chemotaxis, and acute inflammatory response. Moreover, we deducted an elevated IFN score from the altered gene expression in the microglia of *Diora-2* deficient mice. The elevated IFN in microglia may cause or result from infection, inflammation, neurodegeneration, morphologically changes in ramification, or production of reactive oxygen species (ROS) [106-110]. The generation of a conditional knock-out mouse strain that lacks *Diora-2* expression specifically in microglial cells may tackle the questions if the observed phenotype is intrinsic to the cells or not.

Overall, the observed phenotype is less prominent as in *Diora-1* KO mice. In microglia of *Diora-2* deficient mice, we saw alterations in surface marker expression, IFN score, and pathways connected to immune regulation, migration, and intercellular communication.

4.8 DIORA-2 acts in proximity to MRCKA, MRCKB, and TXNL1

To understand the molecular function of DIORA-2 we established a proximity labeling and identified three potential interaction partners, namely MRCKA (CDC42BPA), MRCKB (CDC42BPB), and TXNL1 (Paper IV, Figure 4). Many more hits were identified in each run (Table 1); however, these three hits were the ones shared among all three runs.

On the one hand, TXNL1 is a redox-active cofactor of the 26 S proteasome, widely expressed, and abundant in cytosol and nucleus [111]. On the other hand, MRCK kinases localize to acto-myosin clusters, tightly interact with CDC42, and regulate acto-myosin contractility, thus consequently modulate cell adhesion, motility, and endosomal trafficking [112-114]. Currently, no *Cdc42bpa* or *Cdc42bpb* knock-out mice are available. Hence it is pure speculation if they would share similar phenotypic features to our *Diora-2* mice. However, the knock-out of the MRCK homolog *Gek* in drosophila results in abnormal accumulation of F-actin [115]. Therefore, we should check the actin localization pattern in relevant cells of our *Diora-2* deficient mice.

In neutrophils, Cdc42 has been suggested to regulate cell polarity and migration towards the site of inflammation by crosstalk between CD11b, WASp, and microtubules [116, 117]. Furthermore, CD11b dysregulation has been shown on neutrophils of WASp mutants and Cdc42 KO ([118], unpublished Keszei et al). Based on this connection between Cdc42 and CD11b, it is not surprising that we see altered levels of CD11b on the microglia of *Diora-2* deficient mice. Further, this explains that the gene expression of pathways responsible for chemotaxis, cell migration, and cell-cell adhesion was altered in microglia. In order to understand the impact of DIORA-2 on chemotaxis, cell motility, and polarity, it would be interesting if DIORA-2 knock-out or knock-down cells exhibit impairments in cell migration experiments.

4.9 DIORA-1 and DIORA-2 differ in function but potentially share interaction partners

The two knock-out models established to study the function of the *Diora-1* and *Diora-2* exhibit different phenotypes, however, both show alterations in the immune system. Furthermore, when we identified proteins proximal to DIORA-2, this list overlapped with the BioPlex interaction dataset for DIORA-1 in MRCKA and MRCKB. Considering that DIORA-1 and DIORA-2 potentially share some interaction partners, one could speculate, if they have similar functions, which would go together with the high homology between both proteins. However, this function may be context dependent considering their distinct expression profiles. A double knock-out mouse model could help to understand if both genes complement each other or function independently.

MRCKA and MRCKB interact with Cdc42, which modulates the cytoskeleton via WASp. In dendritic cells, antigen uptake has been shown to be dependent on Cdc42 [119]. When B cells

are specifically deprived of Cdc42 expression, these B cells exhibit a block in development, impaired BCR signaling, antigen presentation, T cell interaction, and fail to differentiate into antibody-producing cells [120, 121]. Another interesting interaction partner for MRCKA is FcγR1A, according to the BioPlex database. FcγRs have been associated with several autoimmune diseases [122] and function both in innate as well as adaptive immune responses, for example, antigen presentation and phagocytosis of IgG-coated particles. To conclude, these cascades may explain how Diora-1 deficiency results in an altered B cell phenotype.

Another member of the MRCK family, namely MRCKG, which also contributes to the reorganization of the cytoskeleton, is associated with chronic kidney disease [123]. Considering that DIORA-1 was also linked to MRCKG in the BioPlex database, this may relate to the kidney phenotype we observed in the knock-out model.

Generally, kinases are promising drug targets. First MRCK inhibitors have been already developed and proved to affect cell motility [124]. That makes this pathway a promising candidate to further explore concerning its relevance in both cancer and immunity as potential targets for novel therapies.

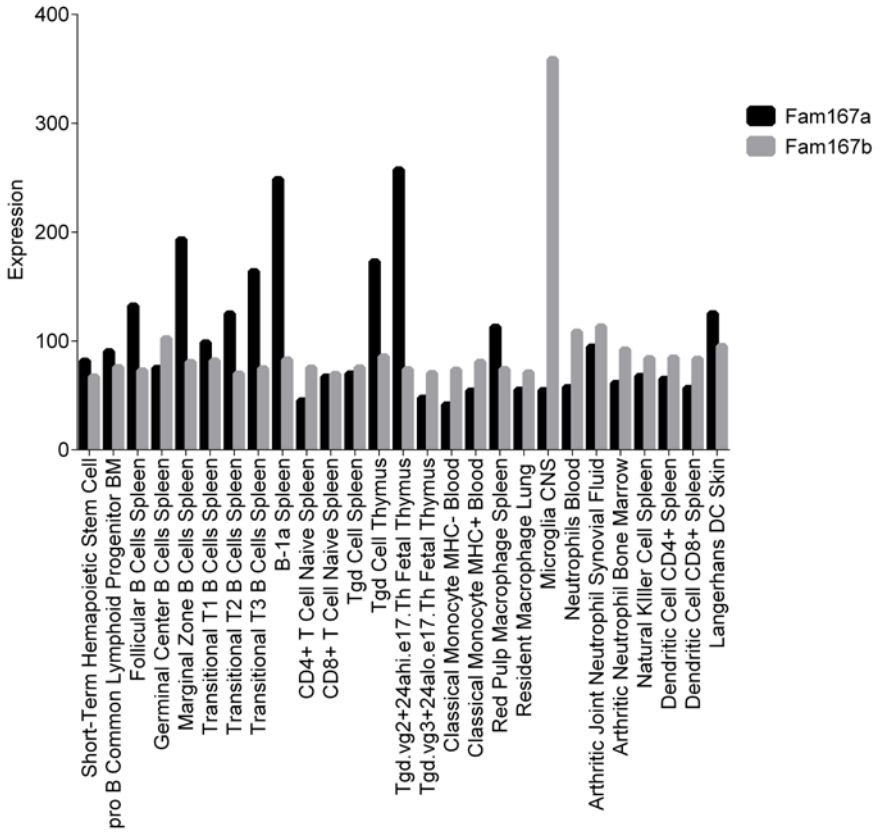

Figure 2: *Fam167a* and *Fam167b* transcription in murine immune cells. Data extracted from ImmGen database.

Hit	Full gene name	Ratio
CDC42BPB	CDC42 binding protein kinase beta (DMPK-like)	6,64
FAM167B	family with sequence similarity 167, member B	6,64
TXNL1	thioredoxin-like 1	6,64
NCCRP1	non-specific cytotoxic cell receptor protein 1 homolog (zebrafish)	6,64
CAT	catalase	6,64
TGM1	transglutaminase 1	6,64
PKP1	plakophilin 1	6,64
S100A11	S100 calcium binding protein A11	6,64
TIMP3	TIMP metallopeptidase inhibitor 3	6,64
MDK	midkine (neurite growth-promoting factor 2)	6,64
CDSN	corneodesmosin	6,64
SERPINB12	serpin peptidase inhibitor, clade B (ovalbumin), member 12	6,64
BLMH	bleomycin hydrolase	6,64
PIP	prolactin-induced protein	6,64
MTX3	metaxin 3	6,64
S100A14	S100 calcium binding protein A14	6,64
XP32	Skin-specific protein 32	6,64
TADA2B	transcriptional adaptor 2B	6,64
POF1B	premature ovarian failure, 1B	6,64
PHLPP2	PH domain and leucine rich repeat protein phosphatase 2	6,64
ZNF829	Isoform 3 of Zinc finger protein 829	6,64
FABP5	fatty acid binding protein 5 (psoriasis-associated)	6,64
MOCOS	molybdenum cofactor sulfurase	6,64
SERPINA3	serpin peptidase inhibitor, clade A, member 3	6,64
AZGP1	alpha-2-glycoprotein 1, zinc-binding	6,64
CDC42BPA	CDC42 binding protein kinase alpha (DMPK-like)	5,38
HAL	histidine ammonia-lyase	4,67
UBFD1	ubiquitin family domain containing 1	3,27
KIAA0319L	KIAA0319-like	2,84
FAM114A2	family with sequence similarity 114, member A2	2,55
SBSN	suprabasin	1,11

Table 1: Proteins identified in SILAC BioID2-FAM167B experiment, sorted by the log2 ratio of abundance sample/control, filtered with crapome. Detailed description of the method can be found in Paper IV.

5 CONCLUSION AND FUTURE PERSPECTIVE

Understanding the function of uncharacterized genes

Many genes in the human genome are merely annotated with very little or nothing know about their function. The undertaking of elucidating the function of these numerous genes will require the consistent effort of many researchers. Nevertheless, this exploration is imperative to proceed in our understanding of the complex networks orchestrating the biological processes within in the cell and complex multicellular organism. It is entirely unforeseeable today, where future biological discoveries will lead us, presumably not only to a refined knowledge on the processes within organisms but potentially also to yet unpredictable applications.

Exploring the *FAM167* gene family

Here, we demonstrate the first effort to characterize the *FAM167* gene family. Our interest in this particular gene family originated not only from the passion for a basic understanding of gene function but also that combined evidence (GWAS, eQTL, present in immune cells) make *FAM167A* a strong candidate gene to contribute to the pathogenesis of rheumatic autoimmune disease. The first steps for characterizing *FAM167A* and *FAM167B* were taken both at the cellular and whole organism level. Based on our findings hitherto, we speculate that the genes exert their function in immune pathways and cell motility. Further complementing *in vitro* and *in vivo* studies will be required to deepen our understanding of their function.

Revealing the etiology of autoimmune diseases

We found that the FAM167A staining pattern in the salivary glands of patients with Sjögren's syndrome correlates with focus score, immunoglobulin and autoantibody levels. Understanding how FAM167A promotes the production of auto-antibodies and infiltration of the target organ might shed light on factors that are involved in the pathogenesis of disease. The etiology of AID is complex, and many pathways are involved. However, if we further explore the function of FAM167 proteins in modulating immunological processes, we might gain insight into their contribution to disease and eventually find novel therapeutic targets.

Suggested model of molecular function

We detected that the DIORA proteins co-localize with vesicles and they are expressed in distinct immune cell subsets. Moreover, we show that the DIORA-proteins interact with the MRCK proteins. These are regulators of CDC42, a small GTPase vital to many cellular pathways. CDC42 is mainly known to control cell polarity and subsequently cell motility via CD11b and other proteins. This may explain our finding of increased CD11b expression and altered cell migration pathways in the microglia of *Diora-2* deficient mice.

Also, B cell-specific deletion of Cdc42 has strong phenotypical consequences like impairment in antibody production, plasma cell differentiation, and interaction with T cells. Thus, we speculate that our observation of IgM increase in *Diora-1* deficient mice and the IgG increase in patients with Sjögren's syndrome on the other side may be regulated through the same pathway.

CDC42 has also been shown to regulate endocytosis, and specifically antigen presentation and uptake. Considering that previous reports propose DIORA-2 as a host factor in viral infections, this might be another mean through which the DIORA-protein exert their immune modulatory function.

Besides, we also observed increased CD45 expression on *Diora-2* deficient microglia. This and other tyrosine kinases act through Src kinases on CDC42. Therefore, it would be interesting to investigate whether the gene neighbors DIORA-1/BLK act in similar pathways or even collaborate.

Considering the inflammatory phenotype in *Diora-1* deficient mice, the altered microglia phenotype in *Diora-2* deficient mice, and the interaction with MRCK proteins, we assume that DIORA proteins cause dysregulation of CDC42 resulting in defective cell migration, antigen processing, and antigen production.

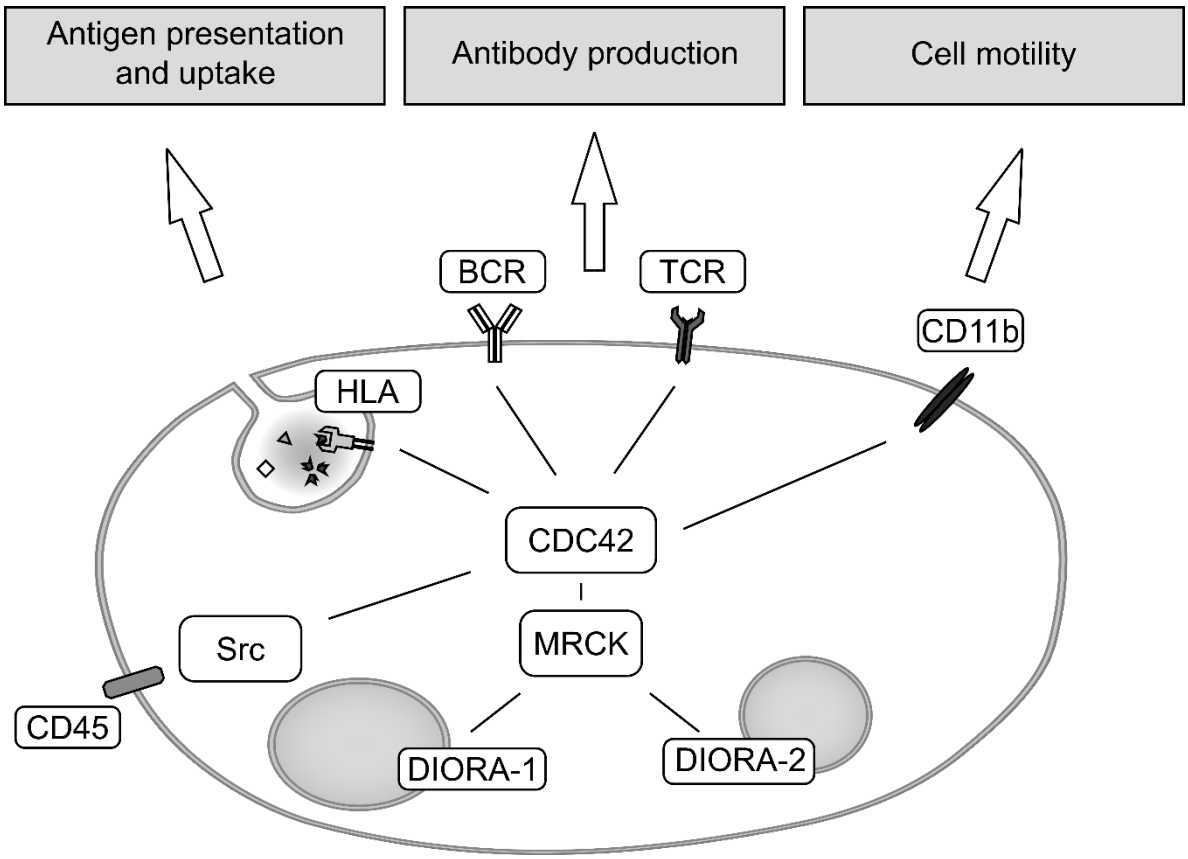

Figure 3: Proposed model on molecular interactions of DIORA-1 and DIORA-2 affecting cell motility and other immune cell functions.

6 ACKNOWLEDGMENTS

I feel sincere gratitude to everyone who contributed or supported me to pursue my PhD education and all the work presented here would not have been achieved otherwise.

First and foremost, I would like to thank my main supervisor *Marie Wahren-Herlenius*. Thank you for reading this thesis so many times. But most of all, thank you for this opportunity to be part of your research team, your support and guidance throughout the journey to become an independent researcher. You have made my time in the lab very inspirational and educational. Also, I feel deep respect for your vision to start from scratch and explore two novel genes.

Alex Espinosa, thank you for being a great co-supervisor and bench neighbor. I have learned a lot from you in the lab and truly enjoyed the scientific discussions we shared.

I also want to take the opportunity to thank the current and former Head of the Rheumatology Unit – *Anca Catrina, Ingrid Lundberg,* and *Lars Klareskog* - for making floor 4 such a great workplace.

To all present and past members of MWH/Espinosa group: Thank you for creating such a supportive work environment! *Guðný Ella*, thank you for introducing me to the lab, your significant contributions to the projects and help with the thesis figures. I truly enjoy the vibe in our office and appreciate that you are always open to discuss questions or having lunch together. *Lauro*, thank you for all your hard work on the projects and the long hours we spent both at the animal house, in the lab and at the FACS machine joining forces to phenotype the *Fam167* KO mice strains. Keep up the great work! *Vijole*, thank you for helping with all those mice experiments, perfusions, serum samples collections, ELISA and genotyping. You are the social glue holding the lab together. *Amina*, thank you for always helping me around the lab and the wonderful IF pictures. You are missed! *Lara A*, thank you for bringing your IHC knowledge to the lab and driving the FAM167A gland project forward. *Susanna*, thank you for initiating the FAM167 project, your continued help, and curiosity for the projects. *William*, thank you for always helping in the lab, keeping the cell lab in good shape and making the BioID experiments work. *Rita*, thank you for sharing your FACS expertise and all the good times we had in and outside the lab. *Albin*, thank you for your efforts in running the clinical studies that have been of value for many of us. *Jorge*, thank you for all your help with the FAM167 projects and the positive spirit in the office and outside. *Marika*, thank you for being our link to the clinic. *Aurélie*, thank you for proofreading our published papers. *Malin*, *Johannes*, *Hanna*, and everyone from the extended group; thank you for contributing to the

positive atmosphere in the MWH lab. To all former members of the group, *Nikolaos*, *Alina, Maria, Sabrina, Kinga, Magda, Lucia, Tian-Lin:* the lab is simply not the same without you. Thank you for all your support and all the best for your careers! Last but not least, good luck to all future FAM167 researchers! There are so many more exciting discoveries to make.

Special thanks are directed to my *co-authors* and *collaborators*. Without our combined efforts there would be no papers or manuscripts to be included in this thesis. *Harald* and *Melanie*, thank you for guiding me through the world of microglia. *Bob Harris* for providing the framework for this research and proof-reading my manuscript. *Johan Grunewald*, thank you for providing as with the BAL cells. *Raoul Kuiper*, thank you for your help with the mouse pathology and your expert opinion. *Susanne Gräslund, Edvard Wigren*, *Elena Ossipova*; thank you for your input on the protein side of these projects and your support in the struggle of protein production. *Robert Schnell*, thank you for showing me your tricks on protein expression.

Thank you, *Lisa Westerberg*, *Maria Sunnerhagen* and *Stephen Malin*, to act as my half-time committee and your valuable suggestions.

I would like to thank all my *colleagues at CMM* for such a friendly and supportive work environment. Thank you, *Agnes,* for our combined effort in organizing the Rheuma-Neuro seminar series. Thank you to all colleagues from the *Neuro* Unit for help with reagents or the FACS machine. Also, special thanks to the *CMM pub crew* members for your contribution to the great atmosphere at CMM.

Stina and *Lillemor*, thank you for helping with administrative tasks. To everyone working at the *IT department*, thank you for solving all computer issues smoothly. To all staff working at the animal facility, especially *Pauline*, thank you for taking good care of our animals.

Micke Karlsson, thank you for hosting me in your lab, your support and everything that I learned about B cells and MARCO. *Kajsa Prokopec*, thank you for being such an inspirational mentor and all the discussions we shared. Thank you, everyone from *LeGroupe/WASP,* who made lab life so productive and inspiring. *Marton Keszei*, thank you for commenting on my preliminary thesis and all the sailing adventures.

I would like to thank everyone who gave advice in times of transition, first of all my mentor, *Vivianne Malmström*. Thank you for your assistance, *Kristina Broliden,* and *Petter Höglund*. Thank you for your advice, *Klas Kärre,* and *Per-Åke Nygren*.

Thank you to my diploma thesis supervisors, *Martin Schwemmle* and *Geoffrey Chase*, for your support in nurturing my interest in research, molecular biology and virology.

Thank you to all my friends from KI: *Shawon*, thank you for your encouragement and discussing science and life with me throughout all these years. *Johanna*, thank you for your support and the photography walks. *Silke,* thank you for always willing to help and spend time together. *Nina,* thank you for all our excursions during our first wonderful winter in Sweden. *Fermín*, I am honored to call you my friend and glad that your collaborations made you revisit Stockholm many times. *Armando*, *Cindy,* and *Cecilia*, thank you for all the great moments we shared both in Stockholm and Mexico. *Petrykk*, *Joanna*, *Milind*, *Jamie*, *Arnika*, *Ivan*, *Marco*, *Anja*, *Anna*, thank you for the good times in and outside the lab. *Cissi* and *Ranjana*, I miss those days where we would discuss science and life with each other every day. You were there for me in difficult times and I am very thankful for your help throughout.

Thank you to all my *friends* - you know who you are - for being patient with me when I was too absorbed by the lab and staying in touch with me regardless. Thank you to all my friends outside of academia: I am always glad to hear your perspective. Thank you *Anja* and *André*, for all the good times spend rowing and ashore.

Thank you, *Ute,* for becoming a member of your family. *Ulrika*, *Micke*, *Lia*, *Michelle*, *Nicole*, *Rebekka*, *Carola*, *Douglas*, *Marcus*, *Vendela*, *Maria*, *Stefan*, *Corinne* and *Laura*, you have all helped me to feel at home in Stockholm.

Thank you to all my *relatives* for your support and always believing in me. Especially, I would like to thank my brother *Thorben*: You are my most important discussion partner.

Very special thanks go to my *parents* who have always encouraged me in my own decisions and assisted me to make them come true.

Many thanks to all those supportive people around me not mentioned here, not because of lack of importance but merely due to lack of space.

Finally, I want to thank all individuals who participated in our clinical studies. It is your curiosity for scientific research that enabled this work.

7 REFERENCES

1. Cooper GS, Stroehla BC. The epidemiology of autoimmune diseases. Autoimmun Rev 2003; **2**:119-25.
2. Klareskog L, Malmstrom V, Lundberg K, Padyukov L, Alfredsson L. Smoking, citrullination and genetic variability in the immunopathogenesis of rheumatoid arthritis. Semin Immunol 2011; **23**:92-8.
3. Ambrosi A, Wahren-Herlenius M. Update on the immunobiology of Sjogren's syndrome. Curr Opin Rheumatol 2015; **27**:468-75.
4. Floreani A, Leung PS, Gershwin ME. Environmental Basis of Autoimmunity. Clin Rev Allergy Immunol 2016; **50**:287-300.
5. Colafrancesco S, Agmon-Levin N, Perricone C, Shoenfeld Y. Unraveling the soul of autoimmune diseases: pathogenesis, diagnosis and treatment adding dowels to the puzzle. Immunol Res 2013; **56**:200-5.
6. Bogdanos DP, Smyk DS, Rigopoulou EI, Mytilinaiou MG, Heneghan MA, Selmi C, Gershwin ME. Twin studies in autoimmune disease: genetics, gender and environment. J Autoimmun 2012; **38**:J156-69.
7. Selmi C, Lu Q, Humble MC. Heritability versus the role of the environment in autoimmunity. J Autoimmun 2012; **39**:249-52.
8. Parks CG, de Souza Espindola Santos A, Barbhaiya M, Costenbader KH. Understanding the role of environmental factors in the development of systemic lupus erythematosus. Best Pract Res Clin Rheumatol 2017; **31**:306-20.
9. Catrina AI, Joshua V, Klareskog L, Malmstrom V. Mechanisms involved in triggering rheumatoid arthritis. Immunol Rev 2016; **269**:162-74.
10. Vojdani A, Pollard KM, Campbell AW. Environmental Triggers and Autoimmunity. Autoimmune Diseases 2014; **2014**:798029.
11. Volkman HE, Stetson DB. The enemy within: endogenous retroelements and autoimmune disease. Nat Immunol 2014; **15**:415-22.
12. Jeffries MA, Sawalha AH. Autoimmune disease in the epigenetic era: how has epigenetics changed our understanding of disease and how can we expect the field to evolve? Expert review of clinical immunology 2015; **11**:45-58.
13. Zhang Z, Zhang R. Epigenetics in autoimmune diseases: Pathogenesis and prospects for therapy. Autoimmunity Reviews 2015; **14**:854-63.
14. Aslani S, Mahmoudi M, Karami J, Jamshidi AR, Malekshahi Z, Nicknam MH. Epigenetic alterations underlying autoimmune diseases. Autoimmunity 2016; **49**:69-83.
15. Melchers F. Checkpoints that control B cell development. J Clin Invest 2015; **125**:2203-10.
16. Koch U, Radtke F. Mechanisms of T cell development and transformation. Annu Rev Cell Dev Biol 2011; **27**:539-62.

17. Suurmond J, Diamond B. Autoantibodies in systemic autoimmune diseases: specificity and pathogenicity. The Journal of Clinical Investigation 2015; **125**:2194-202.

18. Sfriso P, Ghirardello A, Botsios C, Tonon M, Zen M, Bassi N, Bassetto F, Doria A. Infections and autoimmunity: the multifaceted relationship. J Leukoc Biol 2010; **87**:385-95.

19. Rose NR. Negative selection, epitope mimicry and autoimmunity. Curr Opin Immunol 2017; **49**:51-5.

20. Wegner N, Wait R, Sroka A *et al.* Peptidylarginine deiminase from Porphyromonas gingivalis citrullinates human fibrinogen and alpha-enolase: implications for autoimmunity in rheumatoid arthritis. Arthritis Rheum 2010; **62**:2662-72.

21. Klareskog L, Stolt P, Lundberg K *et al.* A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. Arthritis Rheum 2006; **54**:38-46.

22. Vanderlugt CL, Miller SD. Epitope spreading in immune-mediated diseases: implications for immunotherapy. Nat Rev Immunol 2002; **2**:85-95.

23. Lessard CJ, Li H, Adrianto I *et al.* Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren's syndrome. Nat Genet 2013; **45**.

24. Kvarnstrom M, Ottosson V, Nordmark B, Wahren-Herlenius M. Incident cases of primary Sjogren's syndrome during a 5-year period in Stockholm County: a descriptive study of the patients and their characteristics. Scand J Rheumatol 2015; **44**:135-42.

25. Brito-Zeron P, Acar-Denizli N, Zeher M *et al.* Influence of geolocation and ethnicity on the phenotypic expression of primary Sjogren's syndrome at diagnosis in 8310 patients: a cross-sectional study from the Big Data Sjogren Project Consortium. Ann Rheum Dis 2017; **76**:1042-50.

26. Nezos A, Mavragani CP. Contribution of Genetic Factors to Sjogren's Syndrome and Sjogren's Syndrome Related Lymphomagenesis. J Immunol Res 2015; **2015**:754825.

27. Nocturne G, Mariette X. Sjogren Syndrome-associated lymphomas: an update on pathogenesis and management. Br J Haematol 2015; **168**:317-27.

28. Bohnhorst J, Bjorgan MB, Thoen JE, Natvig JB, Thompson KM. Bm1-bm5 classification of peripheral blood b cells reveals circulating germinal center founder cells in healthy individuals and disturbance in the B cell subpopulations in patients with primary sjogren's syndrome. J Immunol 2001; **167**:3610-8.

29. Bohnhorst JO, Thoen JE, Natvig JB, Thompson KM. Significantly depressed percentage of CD27+ (memory) B cells among peripheral blood B cells in patients with primary Sjogren's syndrome. Scand J Immunol 2001; **54**:421-7.

30. Aqrawi LA, Ivanchenko M, Bjork A *et al.* Diminished CXCR5 expression in peripheral blood of patients with Sjogren's syndrome may relate to both genotype and salivary gland homing. Clin Exp Immunol 2018; **192**:259-70.

31. Tengnér P, Halse A-K, Haga H-J, Jonsson R, Wahren-Herlenius M. Detection of anti-Ro/SSA and anti-La/SSB autoantibody-producing cells in salivary glands from patients with Sjögren's syndrome. Arthritis Rheum 1998; **41**:2238-48.

32. Szyszko EA, Brokstad KA, Oijordsbakken G, Jonsson MV, Jonsson R, Skarstein K. Salivary glands of primary Sjogren's syndrome patients express factors vital for plasma cell survival. Arthritis Res Ther 2011; **13**:R2.

33. Hansen A, Lipsky PE, Dorner T. B cells in Sjogren's syndrome: indications for disturbed selection and differentiation in ectopic lymphoid tissue. Arthritis Res Ther 2007; **9**:218.

34. Harris VM, Sharma R, Cavett J *et al.* Klinefelter's syndrome (47,XXY) is in excess among men with Sjogren's syndrome. Clin Immunol 2016; **168**:25-9.

35. Czerwinski S, Mostafa S, Rowan VS, Azzarolo AM. Time course of cytokine upregulation in the lacrimal gland and presence of autoantibodies in a predisposed mouse model of Sjögren's Syndrome: the influence of sex hormones and genetic background. Experimental eye research 2014; **128**:15-22.

36. Khan D, Ansar Ahmed S. The Immune System Is a Natural Target for Estrogen Action: Opposing Effects of Estrogen in Two Prototypical Autoimmune Diseases. Front Immunol 2015; **6**:635.

37. Nordmark G, Kristjansdottir G, Theander E *et al.* Additive effects of the major risk alleles of IRF5 and STAT4 in primary Sjogren's syndrome. Genes Immun 2009; **10**:68-76.

38. Nordmark G, Kristjansdottir G, Theander E *et al.* Association of EBF1, FAM167A(C8orf13)-BLK and TNFSF4 gene variants with primary Sjogren's syndrome. Genes Immun 2011; **12**:100-9.

39. Korman BD, Alba MI, Le JM *et al.* Variant form of STAT4 is associated with primary Sjogren's syndrome. Genes Immun 2008; **9**:267-70.

40. Miceli-Richard C, Comets E, Loiseau P, Puechal X, Hachulla E, Mariette X. Association of an IRF5 gene functional polymorphism with Sjogren's syndrome. Arthritis Rheum 2007; **56**:3989-94.

41. Tzioufas AG, Wassmuth R, Dafni UG *et al.* Clinical, immunological, and immunogenetic aspects of autoantibody production against Ro/SSA, La/SSB and their linear epitopes in primary Sjogren's syndrome (pSS): a European multicentre study. Ann Rheum Dis 2002; **61**:398-404.

42. Fernando MM, Stevens CR, Walsh EC, De Jager PL, Goyette P, Plenge RM, Vyse TJ, Rioux JD. Defining the role of the MHC in autoimmunity: a review and pooled analysis. PLoS Genet 2008; **4**:e1000024.

43. Alvaro-Benito M, Morrison E, Wieczorek M, Sticht J, Freund C. Human leukocyte Antigen-DM polymorphisms in autoimmune diseases. Open Biol 2016; **6**.

44. Samuelson EM, Laird RM, Maue AC, Rochford R, Hayes SM. Blk haploinsufficiency impairs the development, but enhances the functional responses, of MZ B cells. Immunol Cell Biol 2012; **90**:620-9.

45. Simpfendorfer KR, Armstead BE, Shih A *et al.* Autoimmune disease-associated haplotypes of BLK exhibit lowered thresholds for B cell activation and expansion of Ig class-switched B cells. Arthritis Rheumatol 2015; **67**:2866-76.

46. Gourh P, Agarwal SK, Martin E *et al.* Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. J Autoimmun 2010; **34**:155-62.

47. Texido G, Su IH, Mecklenbrauker I, Saijo K, Malek SN, Desiderio S, Rajewsky K, Tarakhovsky A. The B-cell-specific Src-family kinase Blk is dispensable for B-cell development and activation. Mol Cell Biol 2000; **20**:1227-33.

48. Saijo K, Schmedt C, Su IH *et al.* Essential role of Src-family protein tyrosine kinases in NF-kappaB activation during B cell development. Nat Immunol 2003; **4**:274-9.

49. Hom G, Graham RR, Modrek B *et al.* Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med 2008; **358**:900-9.

50. International Consortium for Systemic Lupus Erythematosus G, Harley JB, Alarcon-Riquelme ME *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet 2008; **40**:204-10.

51. Bentham J, Morris DL, Graham DSC *et al.* Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat Genet 2015; **47**:1457-64.

52. Han JW, Zheng HF, Cui Y *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet 2009; **41**:1234-7.

53. Graham RR, Cotsapas C, Davies L *et al.* Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nat Genet 2008; **40**:1059-61.

54. Lee YH, Bae SC, Choi SJ, Ji JD, Song GG. Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. Mol Biol Rep 2012; **39**:10627-35.

55. Gregersen PK, Amos CI, Lee AT *et al.* REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat Genet 2009; **41**:820-3.

56. Okada Y, Wu D, Trynka G *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014; **506**:376-81.

57. Freudenberg J, Lee HS, Han BG *et al.* Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. Arthritis Rheum 2011; **63**:884-93.

58. Onouchi Y, Ozaki K, Burns JC *et al.* A genome-wide association study identifies three new risk loci for Kawasaki disease. Nat Genet 2012; **44**:517-21.

59. Lou J, Zhong R, Shen N *et al.* Systematic confirmation study of GWAS-identified genetic variants for Kawasaki disease in a Chinese population. Sci Rep 2015; **5**:8194.

60. Lee YC, Kuo HC, Chang JS *et al.* Two new susceptibility loci for Kawasaki disease identified through genome-wide association analysis. Nat Genet 2012; **44**:522-5.

61. Chen S, Wu W, Li J *et al.* Single nucleotide polymorphisms in the FAM167A-BLK gene are associated with polymyositis/dermatomyositis in the Han Chinese population. Immunol Res 2015; **62**:153-62.

62. Coustet B, Dieude P, Guedj M *et al.* C8orf13-BLK is a genetic risk locus for systemic sclerosis and has additive effects with BANK1: results from a large french cohort and meta-analysis. Arthritis Rheum 2011; **63**:2091-6.

63. Ito I, Kawaguchi Y, Kawasaki A *et al.* Association of the FAM167A-BLK region with systemic sclerosis. Arthritis Rheum 2010; **62**:890-5.

64. Yin H, Borghi MO, Delgado-Vega AM, Tincani A, Meroni PL, Alarcon-Riquelme ME. Association of STAT4 and BLK, but not BANK1 or IRF5, with primary antiphospholipid syndrome. Arthritis Rheum 2009; **60**:2468-71.

65. Kennedy RB, Ovsyannikova IG, Pankratz VS, Haralambieva IH, Vierkant RA, Poland GA. Genome-wide analysis of polymorphisms associated with cytokine responses in smallpox vaccine recipients. Hum Genet 2012; **131**:1403-21.

66. Torres O, Palomino-Morales R, Vazquez-Rodriguez TR *et al.* Role of the C8orf13-BLK region in biopsy-proven giant cell arteritis. Hum Immunol 2010; **71**:525-9.

67. Liu Y, Ke X, Kang HY, Wang XQ, Shen Y, Hong SL. Genetic risk of TNFSF4 and FAM167A-BLK polymorphisms in children with asthma and allergic rhinitis in a Han Chinese population. J Asthma 2016; **53**:567-75.

68. Shen Y, Liu Y, Wang XQ, Ke X, Kang HY, Hong SL. Association between TNFSF4 and BLK gene polymorphisms and susceptibility to allergic rhinitis. Mol Med Rep 2017; **16**:3224-32.

69. Appel S, Filter M, Reis A *et al.* Physical and transcriptional map of the critical region for keratolytic winter erythema (KWE) on chromosome 8p22-p23 between D8S550 and D8S1759. Eur J Hum Genet 2002; **10**:17-25.

70. Zhou Y, Li X, Wang G, Li X. Association of FAM167A-BLK rs2736340 Polymorphism with Susceptibility to Autoimmune Diseases: A Meta-Analysis. Immunol Invest 2016; **45**:336-48.

71. Lalani SR, Shaw C, Wang X *et al.* Rare DNA copy number variants in cardiovascular malformations with extracardiac abnormalities. Eur J Hum Genet 2013; **21**:173-81.

72. Namjou B, Ni Y, Harley IT *et al.* The effect of inversion at 8p23 on BLK association with lupus in Caucasian population. PLoS One 2014; **9**:e115614.

73. Thalayasingam N, Nair N, Skelton AJ *et al.* CD4+ and B Lymphocyte Expression Quantitative Traits at Rheumatoid Arthritis Risk Loci in Patients With Untreated Early Arthritis: Implications for Causal Gene Identification. Arthritis Rheumatol 2018; **70**:361-70.

74. Lane MA, Wong SK, Daugherty K, Roskey M, Sousa D, Sainten AC, Macoska J. Nucleotide sequence of a human oncogene active in tumors of secretory epithelium. Nucleic Acids Res 1990; **18**:3068.

75. Islam KB, Rabbani H, Larsson C, Sanders R, Smith CI. Molecular cloning, characterization, and chromosomal localization of a human lymphoid tyrosine kinase related to murine Blk. J Immunol 1995; **154**:1265-72.

76. Konsta OD, Le Dantec C, Charras A, Brooks WH, Arleevskaya MI, Bordron A, Renaudineau Y. An in silico Approach Reveals Associations between Genetic and Epigenetic Factors within Regulatory Elements in B Cells from Primary Sjögren's Syndrome Patients. Frontiers in Immunology 2015; **6**.

77. Miceli-Richard C, Wang-Renault SF, Boudaoud S *et al.* Overlap between differentially methylated DNA regions in blood B lymphocytes and genetic at-risk loci in primary Sjogren's syndrome. Ann Rheum Dis 2016; **75**:933-40.

78. Jeyapalan JN, Doctor GT, Jones TA *et al.* DNA methylation analysis of paediatric low-grade astrocytomas identifies a tumour-specific hypomethylation signature in pilocytic astrocytomas. Acta Neuropathol Commun 2016; **4**:54.

79. Coolen MW, Stirzaker C, Song JZ *et al.* Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. Nat Cell Biol 2010; **12**:235-46.

80. Amarillo IE, Nievera I, Hagan A *et al.* Integrated small copy number variations and epigenome maps of disorders of sex development. Hum Genome Var 2016; **3**:16012.

81. Martinez-Martinez MA, De Juan Romero C, Fernandez V, Cardenas A, Gotz M, Borrell V. A restricted period for formation of outer subventricular zone defined by Cdh1 and Trnp1 levels. Nat Commun 2016; **7**:11812.

82. Civini S, Jin P, Ren J *et al.* Leukemia cells induce changes in human bone marrow stromal cells. J Transl Med 2013; **11**:298.

83. Harris SE, Riggio V, Evenden L *et al.* Age-related gene expression changes, and transcriptome wide association study of physical and cognitive aging traits, in the Lothian Birth Cohort 1936. Aging (Albany NY) 2017; **9**:2489-503.

84. Schweitzer CJ, Jagadish T, Haverland N, Ciborowski P, Belshan M. Proteomic analysis of early HIV-1 nucleoprotein complexes. J Proteome Res 2013; **12**:559-72.

85. Panda D, Das A, Dinh PX *et al.* RNAi screening reveals requirement for host cell secretory pathway in infection by diverse families of negative-strand RNA viruses. Proc Natl Acad Sci U S A 2011; **108**:19036-41.

86. Rambow F, Job B, Petit V *et al.* New Functional Signatures for Understanding Melanoma Biology from Tumor Cell Lineage-Specific Analysis. Cell Rep 2015; **13**:840-53.

87. Riker AI, Enkemann SA, Fodstad O *et al.* The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis. BMC Med Genomics 2008; **1**:13.

88. Priyadarshini M, Layden BT. FFAR3 modulates insulin secretion and global gene expression in mouse islets. Islets 2015; **7**:e1045182.

89. Robinson MB, Deshpande DA, Chou J *et al.* IL-6 trans-signaling increases expression of airways disease genes in airway smooth muscle. Am J Physiol Lung Cell Mol Physiol 2015; **309**:L129-38.

90. Le Guillou S, Sdassi N, Laubier J *et al.* Overexpression of miR-30b in the developing mouse mammary gland causes a lactation defect and delays involution. PLoS One 2012; **7**:e45727.

91. Gharib SA, Dayyat EA, Khalyfa A, Kim J, Clair HB, Kucia M, Gozal D. Intermittent hypoxia mobilizes bone marrow-derived very small embryonic-like stem cells and activates developmental transcriptional programs in mice. Sleep 2010; **33**:1439-46.

92. Wade TD, Gordon S, Medland S, Bulik CM, Heath AC, Montgomery GW, Martin NG. Genetic variants associated with disordered eating. Int J Eat Disord 2013; **46**:594-608.

93. Dickinson ME, Flenniken AM, Ji X *et al.* High-throughput discovery of novel developmental phenotypes. Nature 2016; **537**:508.

94. Kim DI, Jensen SC, Noble KA, Kc B, Roux KH, Motamedchaboki K, Roux KJ. An improved smaller biotin ligase for BioID proximity labeling. Mol Biol Cell 2016; **27**:1188-96.

95. Roux KJ, Kim DI, Burke B, May DG. BioID: A Screen for Protein-Protein Interactions. Curr Protoc Protein Sci 2018; **91**:19.23.1-19.23.15.

96. Fairfax BP, Makino S, Radhakrishnan J *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet 2012; **44**:502-10.

97. Salomonsson S, Jonsson MV, Skarstein K, Brokstad KA, Hjelmstrom P, Wahren-Herlenius M, Jonsson R. Cellular basis of ectopic germinal center formation and autoantibody production in the target organ of patients with Sjogren's syndrome. Arthritis Rheum 2003; **48**:3187-201.

98. van der Lee R, Buljan M, Lang B *et al.* Classification of intrinsically disordered regions and proteins. Chem Rev 2014; **114**:6589-631.

99. Gsponer J, Babu MM. The rules of disorder or why disorder rules. Prog Biophys Mol Biol 2009; **99**:94-103.

100. Shulzhenko N, Morgun A, Hsiao W *et al.* Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut. Nat Med 2011; **17**:1585-93.

101. Yin Y, Choi SC, Xu Z *et al.* Normalization of CD4+ T cell metabolism reverses lupus. Sci Transl Med 2015; **7**:274ra18.

102. Zhang H, Li P, Wu D *et al.* Serum IgG subclasses in autoimmune diseases. Medicine (Baltimore) 2015; **94**:e387.

103. Klingelhofer J, Senolt L, Baslund B *et al.* Up-regulation of metastasis-promoting S100A4 (Mts-1) in rheumatoid arthritis: putative involvement in the pathogenesis of rheumatoid arthritis. Arthritis Rheum 2007; **56**:779-89.

104. Abd Elazeem M, Mohammed R, Abdallah N. Correlation of serum interleukin-10 level with disease activity and severity in systemic lupus erythematosus. Egyptian Rheumatology and Rehabilitation 2018; **45**:25-33.

105. Faridi MH, Khan SQ, Zhao W *et al.* CD11b activation suppresses TLR-dependent inflammation and autoimmunity in systemic lupus erythematosus. J Clin Invest 2017; **127**:1271-83.

106. Capuccini B, Lin J, Talavera-López C, Khan SM, Sodenkamp J, Spaccapelo R, Langhorne J. Transcriptomic profiling of microglia reveals signatures of cell activation and immune response, during experimental cerebral malaria. Scientific Reports 2016; **6**:39258.

107. Mathys H, Adaikkan C, Gao F *et al.* Temporal Tracking of Microglia Activation in Neurodegeneration at Single-Cell Resolution. Cell reports 2017; **21**:366-80.

108. Minten C, Terry R, Deffrasnes C, King NJ, Campbell IL. IFN regulatory factor 8 is a key constitutive determinant of the morphological and molecular properties of microglia in the CNS. PLoS One 2012; **7**:e49851.

109. Drokhlyansky E, Göz Aytürk D, Soh TK, Chrenek R, O'Loughlin E, Madore C, Butovsky O, Cepko CL. The brain parenchyma has a type I interferon response that

can limit virus spread. Proceedings of the National Academy of Sciences 2017; **114**:E95-E104.

110. Spencer NG, Schilling T, Miralles F, Eder C. Mechanisms Underlying Interferon-γ-Induced Priming of Microglial Reactive Oxygen Species Production. PLoS ONE 2016; **11**:e0162497.

111. Andersen KM, Madsen L, Prag S, Johnsen AH, Semple CA, Hendil KB, Hartmann-Petersen R. Thioredoxin Txnl1/TRP32 is a redox-active cofactor of the 26 S proteasome. J Biol Chem 2009; **284**:15246-54.

112. Zhao Z, Manser E. Myotonic dystrophy kinase-related Cdc42-binding kinases (MRCK), the ROCK-like effectors of Cdc42 and Rac1. Small GTPases 2015; **6**:81-8.

113. Martinelli S, Krumbach OHF, Pantaleoni F *et al.* Functional Dysregulation of CDC42 Causes Diverse Developmental Phenotypes. The American Journal of Human Genetics 2018; **102**:309-20.

114. Etienne-Manneville S. Cdc42--the centre of polarity. J Cell Sci 2004; **117**:1291-300.

115. Luo L, Lee T, Tsai L, Tang G, Jan LY, Jan YN. Genghis Khan (Gek) as a putative effector for Drosophila Cdc42 and regulator of actin polymerization. Proc Natl Acad Sci U S A 1997; **94**:12963-8.

116. Kumar S, Xu J, Perkins C, Guo F, Snapper S, Finkelman FD, Zheng Y, Filippi MD. Cdc42 regulates neutrophil migration via crosstalk between WASp, CD11b, and microtubules. Blood 2012; **120**:3563-74.

117. Szczur K, Zheng Y, Filippi MD. The small Rho GTPase Cdc42 regulates neutrophil polarity via CD11b integrin signaling. Blood 2009; **114**:4527-37.

118. Keszei M, Record J, Kritikou JS *et al.* Constitutive activation of WASp in X-linked neutropenia renders neutrophils hyperactive. J Clin Invest 2018; **128**:4115-31.

119. Garrett WS, Chen L-M, Kroschewski R, Ebersold M, Turley S, Trombetta S, Galán JE, Mellman I. Developmental Control of Endocytosis in Dendritic Cells by Cdc42. Cell 2000; **102**:325-34.

120. Burbage M, Keppler SJ, Gasparrini F *et al.* Cdc42 is a key regulator of B cell differentiation and is required for antiviral humoral immunity. J Exp Med 2015; **212**:53-72.

121. Gerasimcik N, Dahlberg CI, Baptista MA, Massaad MJ, Geha RS, Westerberg LS, Severinson E. The Rho GTPase Cdc42 Is Essential for the Activation and Function of Mature B Cells. J Immunol 2015; **194**:4750-8.

122. Bournazos S, Woof JM, Hart SP, Dransfield I. Functional and clinical consequences of Fc receptor polymorphic and copy number variants. Clin Exp Immunol 2009; **157**:244-54.

123. Yasukochi Y, Sakuma J, Takeuchi I, Kato K, Oguri M, Fujimaki T, Horibe H, Yamada Y. Identification of CDC42BPG as a novel susceptibility locus for hyperuricemia in a Japanese population. Mol Genet Genomics 2018; **293**:371-9.

124. Unbekandt M, Croft DR, Crighton D *et al.* A novel small-molecule MRCK inhibitor blocks cancer cell invasion. Cell Commun Signal 2014; **12**:54.